NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

ANALYZING SUCCESS OF NAVY ENLISTEES WITH MORAL WAIVERS

by
Lyle D. Hall

September 1999

Thesis Advisor:	Samuel E. Buttrey
Second Reader:	Robert A. Koyak

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 1999	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE ANALYZING SUCCESS OF NAVY ENLISTEES WITH MORAL WAIVERS	5. FUNDING NUMBERS
6. AUTHOR(S) Hall, Lyle D.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Commander, Navy Recruiting Command Arlington, VA	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

Unsuitable attrition of recruits from the Navy is a costly problem. This thesis compares unsuitable attrition rates for recruits with moral waivers to the rates of recruits without moral waivers. Unsuitable attrition is also modeled using both logistic regression and classification trees for the recruits who received moral waivers. The comparison and models were completed on two data sets, one that contained all recruits for FY's 95-96 and a subset of the data modified to account for a known bias in the data. The comparison of unsuitable attrition rates found that recruits with moral waivers do have a significantly higher rate of unsuitable attrition than that of recruits without moral waivers. The prediction models produce "significant" variables, but they predict poorly when applied to the data. However, it is found that recruits who are not high school graduates and receive a moral waiver are the most likely unsuitable attrition losses. Unsuitable attrition rates differ when the data collection error is addressed, but both data sets result in the same conclusion that recruits with moral waivers have a higher unsuitable attrition rate than recruits without moral waivers.

14. SUBJECT TERMS Unsuitability Attrition, Moral Waivers, Recruiting, Logistic Regression, Classification Tree	15. NUMBER OF PAGES 110
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFI- CATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500 Standard Form 298

Approved for public release; distribution is unlimited.

ANALYZING SUCCESS OF NAVY ENLISTEES WITH MORAL WAIVERS

Lyle D. Hall
Lieutenant, United States Navy
B.S., Texas A&M University, 1991

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1999

Author: Lyle D. Hall

Lyle D. Hall

Approved by: Samuel E. Buttrey

Samuel E. Buttrey, Thesis Advisor

Robert A. Koyak

Robert A. Koyak, Second Reader

Richard E. Rosenthal

Richard E. Rosenthal, Chairman
Department of Operations Research

ABSTRACT

Unsuitable attrition of recruits from the Navy is a costly problem. This thesis compares unsuitable attrition rates for recruits with moral waivers to the rates of recruits without moral waivers. Unsuitable attrition is also modeled using both logistic regression and classification trees for the recruits who received moral waivers. The comparison and models were completed on two data sets, one that contained all recruits for FY's 95-96 and a subset of the data modified to account for a known bias in the data. The comparison of unsuitable attrition rates found that recruits with moral waivers do have a significantly higher rate of unsuitable attrition than that of recruits without moral waivers. The prediction models produce "significant" variables, but they predict poorly when applied to the data. However, it is found that recruits who are not high school graduates and receive a moral waiver are the most likely unsuitable attrition losses. Unsuitable attrition rates differ when the data collection error is addressed, but both data sets result in the same conclusion that recruits with moral waivers have a higher unsuitable attrition rate than recruits without moral waivers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The selection of qualified enlistees who are likely to succeed and provide value to the Navy is a continuing problem. However, the number of fully qualified enlistees does not meet recruiting requirements. Therefore, the Navy has a system in place that allows waivers to individuals who do not meet certain pre-selection requirements. The use of such waivers allows the Navy to meet enlistment numbers even when enough applicants who meet pre-screening requirements are not available. In particular, there is a waiver in place known as a "moral waiver" for individuals with a background such as drug use or other types of criminal behavior, which brings their morality into question.

Due to the high cost of training a recruit, it is imperative that the Navy select individuals who will complete their service and provide the fleet a benefit. To help with the selection process, the Navy needs a way to determine an enlistee's chance of success even when a waiver is required. With respect to moral waivers, there are two important issues. First, do individuals who are allowed into the Navy under moral waivers have a higher unsuitable attrition rate than those who are not? Also, as a follow-on, are there any specific identifiable characteristics of enlistees with moral waivers who attrite before completion of their enlistment that can be used as selection criteria?

To analyze this problem, Navy Recruiting Command provided enlisted accessions data for fiscal years 95 and 96. The data fields include whether the individual had a moral waiver, whether they attrited in their first two years of active duty, and other identifying characteristics of each enlistee. The initial data set consisted of 96,843 records.

This data set was used to create two data sets for this study. The first was the entire data set that was provided. The second data set was a subset of the entire data set. This subset was created based on a known recording bias in the data caused by program waivers. Certain rates and/or enlistment programs within the Navy require higher entrance standards than others. If an individual required a waiver for a specific rate/program (a program waiver), but did not require a waiver for enlistment, he or she was still recorded in the waiver group. To attempt to account for this situation, the data for the rates or programs were removed since the specifics of the waivers were undeterminable. The remaining records constitute the second data set.

It was found that recruits with moral waivers do have a significantly higher unsuitable attrition rate than that of recruits without moral waivers. In the entire data set, recruits with moral waivers (34.02%) had a 9.3 percent higher unsuitable attrition than that of recruits without moral waivers (24.70%). It was found in the modified data set that recruits with moral waivers (37.26%) were 9.9 percent more likely to have unsuitable attrition than recruits without moral waivers (26.34%).

Prediction models were created by modeling unsuitable attrition using both logistic regression and classification trees for the recruits who received moral waivers. This was undertaken on both data sets to identify characteristics of recruits that could be used to predict their success/failure. It was found that the EDCERT code of N (not a high school graduate) has significance in each of the models. "Non-grad" is one of three possible codes in the EDCERT data referring to high school education status: graduate, G.E.D., and non-grad. The code was found as a high loss probability coefficient in the

logistic models and it was combined with other characteristics to predict loss in the tree models. However, the additional characteristics associated with an unsuccessful non-graduate differ between the two data sets. One other important effect was noted in all the models. This was that recruits with a Race/Ethnic code of Asian had greatly reduced probabilities of unsuitable attrition.

These models did produce "significant" prediction variables, but when they were tested with the given data set there was no substantial prediction capability found in any of the models. Therefore, it is not recommended that recruits be excluded based on these models. However, it is recommended that the selection of non-graduate recruits be analyzed closely. This is suggested since there has been a change of policy at the time of this thesis that allows more high school non-graduates to enlist. This study identifies them as a group with higher unsuitable attrition probability when they have a moral waiver, which raises a concern about the effect of this policy on future unsuitable attrition rates.

I. INTRODUCTION

A. PROBLEM IDENTIFICATION

With the U.S. Navy maintaining an all-volunteer force, the selection of qualified enlistees who are likely to succeed and provide value to the Navy is a critical problem. With the low civilian unemployment rates that are present at the time of this thesis, the Navy has recently been unable to meet recruiting goals. Starting in FY99, the Navy began allowing a higher percentage of high school non-graduates to be enlisted as one step in attempting to meet recruiting goals. However, even under previous recruiting standards, there were ways that individuals who did not meet standards could enter the Navy. Individuals who do not meet all of the basic standards can be granted waivers in order to enlist in the Navy.

Waivers can be granted for a broad range of reasons. The use of waivers allows the Navy to look at candidates more carefully who have characteristics that may effect their ability to perform successfully in the Navy. Applicants who require a waiver for any enlistment eligibility are only processed if they are considered to be a particularly desirable candidate. Waivers can be granted for age, number of dependents, mental qualifications, moral qualifications, medical qualifications, education, and for other reasons. Waiver types will be discussed in more detail in the data/methodology section.

Due to the high cost of training a recruit, it is imperative that the Navy select individuals who will complete their service and provide a benefit to the Navy. To help with the selection process, the Navy needs to determine an enlistee's chance of success.

The ability to choose recruits who will succeed is also important in getting quality recruits to initial assignments and who will finish their initial enlistment period. Bohn and Schmitz (1996) found that 26% of recruits for FY 92-93 attrited before finishing their first two years of active service. This high rate of attrition not only wastes valuable training dollars, but also decreases the number of trained sailors available to fill advanced assignments.

Recent high attrition rates in the Navy have raised many questions concerning their underlying cause. Many fleet commanders have stated concern that disciplinary problems and attrition can be related to individuals who received moral waivers. This study will look at moral waivers to examine the validity of their concern.

Bohn and Schmitz (1996) found that between 16.3% and 21.0% of the recruits for FY's 92-96 required moral waivers each year. There was a total of 43,948 recruits entering with moral waivers out of a total of 247,368 recruits, or a percentage of 17.8% over the five-year period. With a substantial percentage of recruits receiving moral waivers, it is of interest to determine if they do have an identifiably higher attrition rate.

B. WHAT IS A MORAL WAIVER?

To study this problem, it is important to understand what constitutes a moral waiver. A moral waiver is an exemption from Navy enlistment standards granted for the following reasons: civil offenses, drug abuse, and alcohol abuse. It must be noted that the policy as of Dec. 1998 for granting moral waivers (Enlisted Policy Gram 27-98 with change 01-99) is not the same policy that was in effect for the recruits in the data set for this study. The current policy replaced the chapter on moral waivers in

COMNAVCRUITCOMINST 1130.8E (Navy Recruit Manual). All policies discussed in this section will be from COMNAVCRUITMANINST 1130.8D with change 31 incorporated. This is the Navy Recruit Manual that was in effect during FY 95-96 which corresponds to the data used for this study.

In the case of a civil infraction, a waiver is only required for offenses where there was a conviction, adverse adjucation or which were processed through a pre-trial intervention program. If an applicant had infractions in more than one category that required a waiver, he or she is given a waiver for the most serious offense. Several violations at the same time and place are counted as a single transgression. The recruiter at the local recruiting command makes the determination as to whether an individual requires a waiver. An applicant who exceeds the limits for enlistment can still be enlisted with a waiver and the approval of Commander, Navy Recruiting Command. This is done if it is clear that granting the exception to the waiver policy is in the best interest of the Navy. Table 1 shows the waiver policy for civil offenses.

A list of offenses and their classification, as specified in the Navy Recruit Manual, is provided in Appendix A. The classifications used in the Navy Recruit Manual take precedence over state law classifications except in the case where a state classifies a crime as a felony. If a crime is classified as a felony by the state, it is considered a felony for enlistment purposes. Additionally, the "Navy Sunset Rule" overrides the requirement for waivers in the case of some Minor Non-Traffic/Minor Misdemeanors and Non-Minor Misdemeanors. The Navy Sunset Rule decision flow chart is included as Figure 1. If the rule applies, a waiver is not required.

Table 1: Waiver Policy for Civil Offenses

Offense	Number of Offenses	Waiver Authority
Minor Traffic Violations	6 or more in a 12 month period prior to DEP-IN. 10 or more within 3 years prior to DEP-IN	CO, NAVCRUITDIST
Minor Non-Traffic/Minor Misdemeanors	3-5	CO, NAVCRUITDIST
	6 or more	No waiver authorized
Non-Minor Misdemeanor	1-2	CO, NAVCRUITDIST
	3 or more	No Waiver authorized
Felonies	1 or more	Commander, NAVCRUITCOM
	Juvenile Felony (Note 1)	CO, NAVCRUITDIST

Source: Navy Recruit Manual

Note (1): A single felony before age 14, 3 or more years ago without alcohol, drugs, or physical violence and no other charges except minor traffic violations.

The Navy Sunset Rule has the following restrictions:

1. If the Navy Sunset Rule can not be applied to ALL civil waivers it can not be applied. Civil waivers must be conducted for all convictions.

2. The Navy Sunset Rule can not be used to eliminate the need for an alcohol abuse waiver although it may be used to cover civil waivers if an alcohol abuse waiver is required. Alcohol abuse waivers are not considered civil waivers.

Alcohol and drug abuse waivers operate similar to civil waivers. Native American applicants who have used peyote for religious purposes do not require a waiver for that use. However, they must be notified that the use of peyote is not allowed while in the delayed entry program or on active duty. Table 2 provides a list of drug/alcohol offenses and the waivers required.

Figure 1: Navy Sunset Rule Decision Flow Chart

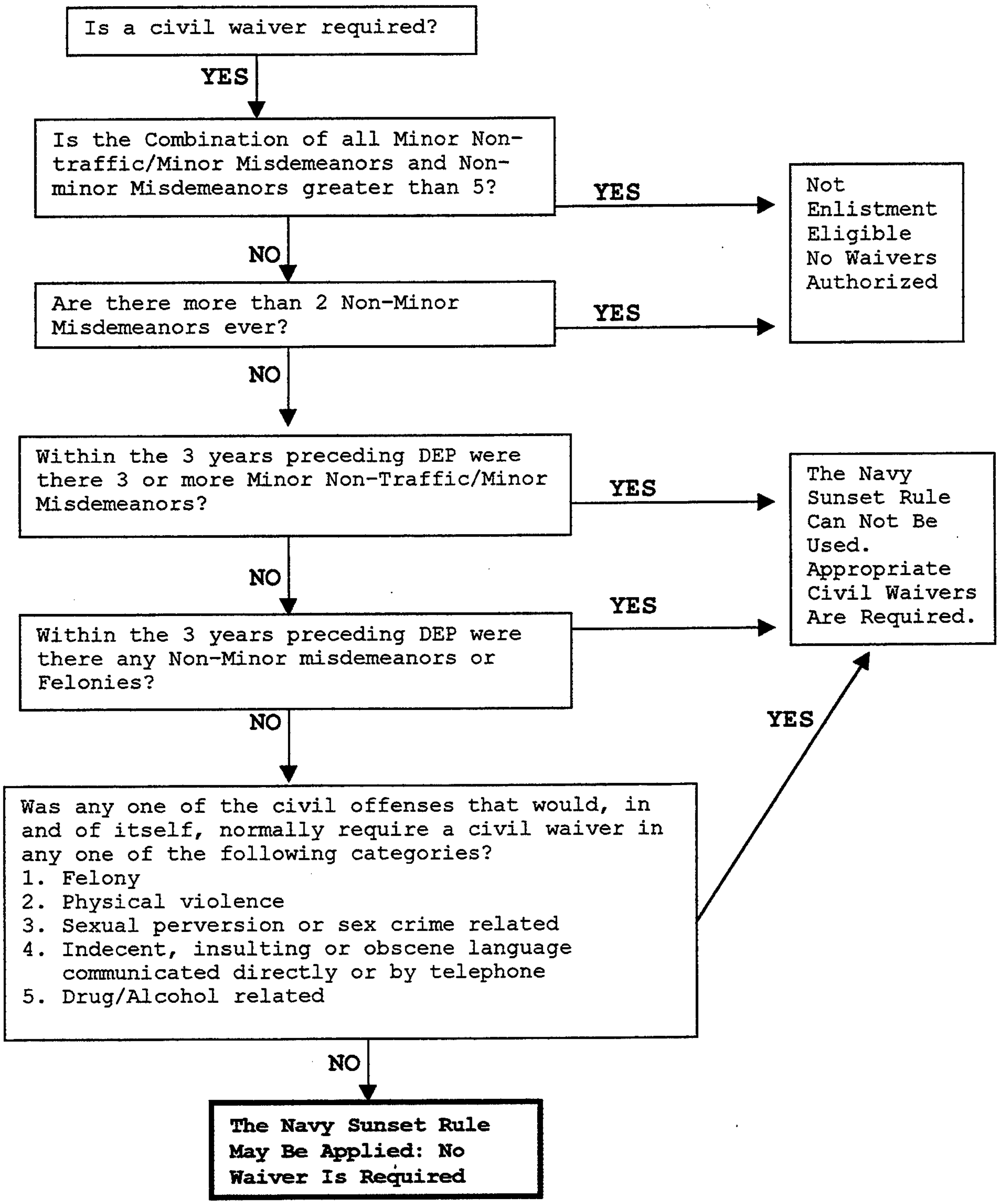


Table 2: Waiver Policy for Alcohol and Drug Abusers

Alcohol/Drug Abuse	Waiver Authority
Experimental/casual use of Marijuana	No waiver required
Convicted of drug abuse or single alcohol related offense	Appropriate civil waiver authority
Convicted of 2 or more alcohol related offenses (except 2 behind the wheel offenses)	Appropriate civil waiver authority CO, NAVCRUITDIST for alcohol/drug waiver
2 behind the wheel offenses	Commander, Navy Recruiting Command
Prior psychological or physical dependence upon any drug or alcohol	Commander, Navy Recruiting Command
Abuse of Narcotics, Hallucinogenic, or Psychedelic drugs within one year	No waiver authorized
Abuse of Narcotics, Hallucinogenic, or Psychedelic drugs over one year ago	CO, NAVCRUITDIST
Abuse of Stimulant or Depressant drugs within the past six months	No waiver authorized
Abuse of Stimulant or Depressant drugs between six months and one year ago	CO, NAVCRUITDIST
Abuse of Stimulant or Depressant drugs over one year ago	No waiver required
Any drug/alcohol abuse while in DEP	NOTE (1)
Drug trafficking/supplying	No waiver authorized

Source: Navy Recruit Manual

Note(1): Interview by the NAVCRUITDIST Commanding Officer and waiver, if required. No recruit can go to training command who has used marijuana in the last 30 days.

There are numerous categories of waivers and a recruit can receive both civil and alcohol/drug waivers. However, only the worst of these is included in data recording. The section on moral waivers also includes requirements for program waivers, which are waivers for specific programs/ratings. Although these waivers will not be used for analysis, their implication for the study will be discussed in the data/methodology section.

As noted above, the waiver policy underwent changes from the time the data was collected to the initiation of this study. There are numerous minor changes, though only a few major ones. First, the Navy Sunset Rule was dropped, as were special rules for juvenile felonies. Second, program waivers for civil offenses were moved to another chapter to be treated separately from standard moral waivers. Finally, and most significantly, a mandatory waiting period after an adverse alcohol/drug adjudication was added to restrict entry into the delayed entry program. The final model will be analyzed to determine if these changes were consistent with the recommendations of the analysis.

C. PREVIOUS STUDIES

A small number of studies have been conducted on the effects of moral waivers on performance/attrition. These include studies by the recruiting command and previous Naval Postgraduate School thesis projects. These studies agreed that among enlistees with certain classes of moral waivers, there were higher attrition rates, but the magnitude of effects differed. The reliability of the data used for most moral waiver studies was brought into question in one study. I will summarize the results of four studies and one article for background and use in conclusions and recommendations.

Bohn and Schmitz (1996) conducted a study using a 20% sample of FY 92-93 accessions to compare the difference in attrition rates for moral waiver enlistees and non-moral waiver enlistees. The authors conducted a regression analysis to identify variables that are predictors of attrition. They found that recruits with waivers for criminal behavior had five percent more attrition (over two years) then those without. However, the authors

also found that the attrition rate of those receiving non-criminal, drug or alcohol abuse moral waivers was not significantly different than those without waivers. In their model they used the entire 20% sample and found seven variables that had a greater predictive effect on attrition than did the presence or absence of a moral waiver. Overall, they found that the effects of moral waivers were not uniform over gender and education groups. They determined that eliminating certain combinations of gender or education groups that required criminal waivers would not be cost-effective, comparing the projected attrition savings to the increased cost of recruitment to replace the enlistees.

Bohn (1998) conducted a case study of sailors from the U.S.S. Eisenhower who left the Navy during or at the end of their first term between FY 91 and 3rd quarter FY 97. He found that those with moral waivers had a 31.9% chance of being discharged for misconduct. The comparable rate for those without waivers was 23.7%, an eight percent difference. He also pointed out that if criminal waivers were looked at separately; the discharge rate was higher than 35%, while the other categories of moral waivers were similar to the no-waiver category. The study concluded that the cost of changing the moral waiver policy was greater than that of keeping the current policy.

Etcho (1996) conducted a MS thesis study that looked at the effects of moral waivers on unsuitability attrition in the Marine Corps. However, he also included some Navy data in his study. For FY 88-91, Etcho (1996) found that first-term (4-year) attrition rates for recruits without moral waivers ranged from 14.38 to 14.81%, and from 16.35 to 18.5% for recruits with moral waivers in the Marine Corps. When moral waivers were examined by category, the traffic offense group was similar to the non-waiver group

while all others were higher, with the felony group having the highest attrition rate (18.1%-21.9%). In Navy FY 88 accessions, Etcho (1996) found the attrition rate for recruits with no moral waivers to be 18.3%, compared to 25.0% for those with moral waivers. For each of the specific categories examined, the attrition rate was higher for waivers than for non-waivers, with alcohol at 20.31% attrition for those with waivers being the lowest of the categories. Among those with moral waivers in the Marine Corps data, enlistees who had not completed high school had the highest attrition rates. He recommended discontinuing moral waivers, with the exception of traffic violations, for non-high school graduates and to completely discontinue the felony waiver in the Marine Corps.

Connor (1997) studied the effects of pre-service criminal history on performance in the Navy. He used accessions into the Navy from Illinois in years 1981 to 1987 and from Florida in years 1984 and 1988. Connor (1997) found that Florida recruits with felony arrests, felony convictions, and non-felony convictions had an attrition rate that was more than 7 percentage points higher than for those without criminal records. Among those with non-felony arrests, the difference was 4.4 percent higher. For Illinois recruits the effect on attrition rates was larger: 11.9 percent for felony arrests, 12.4 for felony convictions, 8.4 for non-felony arrests, and 6.5 for non-felony convictions. Connor (1997) found that recruits with criminal backgrounds were less likely to be promoted to E-4, less likely to be eligible for re-enlistment, and less likely to remain in the Navy beyond their initial term.

Connor (1997) also found that 97.3% of the convicted juvenile felons in his study were not identified by the moral waiver process. For the Florida group, 91% of adult convicted felons did not have a moral waiver for their criminal violations. This finding brings into question the effectiveness of the moral waiver policy as well as the reliability of the data used for all moral waiver studies.

Kannapel (1998) addressed the concern from fleet commanding officers that the increased number of moral waivers was leading to an increased number of discipline cases. He noted that there is a direct relationship between the number of individuals with moral waivers and discipline cases involving members with moral waivers, as would be expected. It is pointed out that 15% of accessions in 1996 had moral waivers, compared to 13% in 1995. However, these percentages of moral waivers do not match the rates given by Bohn and Schmitz (1996) for FY95 and FY96. It is stated that the number of recruits with moral waivers declined to 13% in 1997 and 11% in 1998, which will result in fewer discipline cases involving individuals with moral waivers. Therefore, he says, there is no need to change waiver policy. However, as stated earlier, the numbers he uses do not appear to be consistent with other studies. This leads to a question about the true percentages. There is also no statement in the article addressing the concern that recruits with moral waivers have a higher incidence of discipline problems. The only statement is that discipline cases involving members with moral waivers will decline as the number of moral waivers declines.

All of these background articles concur that recruits with moral waivers have a higher attrition rate than those without, except Kannapel (1998) who does not address this

claim. It also appears that traffic violation waivers do not bring about a significant difference in attrition. The big question in these articles is the severity of the difference in attrition rates for recruits with moral waivers versus those without.

II. DATA AND METHODOLOGY

A. DATA

Data for this study was provided by Naval Recruiting Command, which merged data from two U.S. Navy data sources: PRIDE and TrainTrack. PRIDE is the Personalized Recruiting for Immediate and Delayed Enlistment database. It is the Navy's reservation system for initial entry that records information about recruits prior to their entry into DEP and while in DEP. TrainTrack is a database that tracks the training pipeline of a sailor throughout his or her career and retains training and career data. The data set consists of FY95 and FY96 enlisted accessions and contains information about their active duty status as of 30 June 1998. There are a total of 96,843 records, with each record identified by the recruit's social security number.

Within each record are 36 characteristics of the recruit. These include variables that will be used for prediction models, identification of enlistment program and rating, waiver data, whether the recruit attrited, and the reason for attrition. Table 3 provides a description of the data fields in each record. Possible entries for unclear variables are included as Appendix B. A sample of the data is included as Appendix C.

Many of the characteristics for each record contain empty ("null") fields due to the type of information provided in them. Fields such as DEPDAYS and NAVYLOSS will necessarily have null fields for recruits who do not enter DEP or were not a loss. However, the fields ATTRITE and ACC_WAIV have specific entries to identify all possibilities.

Table 3: Data Descriptions

Variable	Description
SSN	Recruit's social security number
RESDT_TT	Initial date of entry into DEP (delayed entry program)
CANDATE	Date of accession into the Navy
DEPDAYS	Days spent in DEP
LOSSDATE	Date of leaving active service
NAVYLOSS	Reason for leaving active service
SERVDAYS	Days on active duty
ATTRITE	Two-year attrition code
PRIOR_SV	Did recruit have prior military service?
AFQT	Armed Forces Qualification Test scores
GS, AR, WK, PC, NO, CS, AS, MK, MC, EI	Score of individual sections that make up the Armed Forces Qualification Test
SENGRAD	Education at time of reservation
EDYRS	Years of education at time of accession
CIV_CODE	Detailed education codes
SEX	Male or female
RACE	Race identifier
ETHNIC	Ethnicity codes
DOB	Date of birth
PAY_TT	Pay-grade at last entry in database
PAYGRADE	Accession pay-grade
PROGRAM	Program enlisted for
RATE	Rate enlisted for
TERM	Length of enlistment term
DEPEND	Number of dependents
ACC_WAIV	Accession waiver category
NRD	Recruiting district recruited from

B. DATA ERRORS

Within the data set there is a known pre-existing data recording error. As was mentioned in the background, along with enlistment waivers, there are also waivers for specific programs. Some programs and rates require a waiver that is not required for a

standard enlistment, such as stricter requirements on drug history for rates that require security clearances. When the data was recorded, a recruit was identified as having received a moral waiver even if it was a program waiver and a normal moral waiver was not required. This leads to over-reporting of the number of moral waivers. According to Navy Recruiting Command, this error has been corrected for future data sets, but is inherent to any data that can currently be used for a two-year attrition study.

There are also some fields that contain null entries where data is expected, such as a 0 score for the AQFT. Since this study will be broken into separate sections that use different data, records with null fields are removed only when they affect the particular section of the study that is being conducted. The fields in each section will also be verified against other data cells to check for data errors present in the fields being used.

C. METHODOLOGY

For this study, I will be looking at two-year "unsuitable attrition" in the U.S. Navy, which will be defined by specific Navy loss codes. The goal is to determine if there is a significant difference in unsuitable attrition between recruits who entered with moral waivers and those who did not. I will then identify characteristics of recruits with moral waivers who attrited to be used in the future determination of who should be granted moral waivers. Since two-year attrition is the guide in this study, and the data was collected on 30 June 1998, some of the records do not meet the two-year requirement. In total there are 10,028 records from 4th quarter FY96 that do not meet the two-year requirement and they were removed from the data set.

The remaining data numbers 86,815 records. Using this data, two separate analyses will be conducted. First, I will compare percentages of two-year attrition for recruits who received moral waivers and those who did not. These results will then be analyzed to determine if a significant difference in attrition rates is evident. Second, the records of recruits who received moral waivers will be used to create prediction models of success or failure among recruits granted moral waivers.

Within each of the analyses, two separate data sets will be used. The first will use the data as it was recorded. Then, the second will modify the moral waiver group due to the known data error. All recruits – with or without moral waivers – in programs or rates with program waivers will be removed from the data. This will remove the question of whether their waiver was a moral waiver or a program waiver and just look at recruits that are known to have a moral waiver if a waiver is identified. In making this adjustment, the records of the recruits identified in Table 4 will be removed (programs and rates are defined in Appendix B). Based on the data available this includes all program waiver possibilities, although it may not be exhaustive. However, it is enough of the possibilities to identify if a difference exists from the entire data set. This data set numbers 56,510 records.

Table 4: Program Waiver Possibilities

Programs	Rates					
Nuclear	AC	AW	CTA	CTI	CTM	CTO
Advanced Electronic	CTR	CTT	DS	DT	ET	ETS
Advanced Technical	EW	FC	GM	HM	IS	MMS
Diver	MN	MSS	MT	OS	RM	SKS
JOBS	STG	STS	TM	TMS	YNS	

1. Comparison Model

For the comparison models, the percentage of unsuitable attrition for recruits with moral waivers will be compared against those without moral waivers. For this analysis, the attrition section, accession waiver section, service days, and Navy loss section of the data (as identified in Table 3) will be used from the data set. Attrition rates will be computed for each type of attrition loss, as identified in Navy loss codes, with an overall attrition rate determined for attrition losses and for all losses. Then a rate will be determined for the codes determined as unsuitability attrition for this study. The following codes from the Navy Loss Codes (included in Appendix B) will be considered unsuitability attrition: 817-825, 831-833, 857-873, 881-890, 901-902, 911, 970-972. This will be conducted for the entire data set and the modified data set.

Once percentages are determined, tests for significance of the differences will be conducted on the percentages. This will be conducted by procedures for comparing population proportions (Devore, 1995, pp. 375-377). The test will hypothesize that the proportions are equal, and then test to see if we reject the hypothesis, or if we fail to reject it. This test assumes that the samples are large enough for the usual Normal approximation to hold. The test rules are outlined in Figure 2.

Figure 2: Large Sample Population Proportion Test

Null Hypothesis: $p_1 - p_2 = 0$ where p_1 and p_2 are the population proportions of populations 1 and 2

Alternative Hypothesis: $p_1 - p_2 \neq 0$

Test Statistic: $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{m} + \frac{1}{n})}}$ where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions from populations 1 and 2. $\hat{p}$ is the weighted average of the two population samples and $\hat{q} = 1 - \hat{p}$.

Rejection Region: $z \geq \left| z_{\alpha/2} \right|$

Source: Devore, 1995

2. Prediction Models

This section will develop a model to predict success or failure of recruits with moral waivers. A logistic regression and a classification tree will be developed in S-Plus® 4.0 (release 2) using the records of recruits with moral waivers. To develop the models, the following data items (as identified in Table 3) were used at the start of model development: DEPDAYS, PRIOR_SV, AFQT, EDCERT, SEX, RACE, ETHNIC, DOB, PAYGRADE, PROGRAM, TERM, DEPEND, ACC_WAIV, and NRD. Minor modifications were made to the data to allow more practical use in prediction models. Modifications consisted of combining RACE and ETHNIC into one field, converting DOB to age, and combining NRD into recruiting regions instead of districts, with regions as identified in Appendix B. RACE and ETHNIC were combined so that the RACE code was used except in the case where the ETHNIC code identified the individual as Hispanic, which is not included among the RACE codes. To do this conversion, Hispanic was

added to the RACE codes and any recruit who has any of the Hispanic codes in the ETHNIC section (codes 1, 4, 6, 9, S) were moved to the Hispanic race section.

Finally, models were developed both for the entire data set and for the modified data set. The models used were logistic regression and classification trees. These models are explained separately in the following sections.

a. Logistic (Logit) Regression

The logistic regression model has been widely used for attrition studies. Since the response in the attrition data is a binary variable ("attrite" vs. "not attrite"), a procedure that models binary variables is needed. "The logistic regression model is a generalized linear model (GLM) that is specially designed for modeling binary and more generally binomial data" (Chambers and Hastie, 1992). The logistic regression provides as a result a probability that by definition is bounded by zero and one. The probability corresponds to the chance of attrition of an individual based on his or her characteristics. In particular, the logistic model is

$$\hat{p} = 1 / (1 + \exp(-X\beta))$$

where $\hat{p}$ is the computed probability of attrition for a recruit, X is the vector of characteristics of a recruit and β represents the vector of regression coefficients for the given characteristics (Hamilton, 1992). A model is chosen for prediction by removing characteristics that do not appear to affect the response variable at some pre-determined level of confidence. This is continued in an iterative process until all of the characteristics are significant to the pre-determined level of confidence.

b. Classification Tree

A classification tree is another alternative for binary data. Construction of a tree is a recursive process that looks one step ahead. In this process it looks to maximize the reduction in deviance in a single split, without looking at the entire tree. Deviance is defined in one node as:

$$\text{Deviance} = -2 * \sum_k n_k * \log(p_k)$$

where k indexes the classes in the node ("attrite" and "not attrite"), n_k represents the number of cases of class k in the node and p_k is the observed proportion of class k in the node. A choice is made so as to maximize the reduction in deviance. The two new nodes ("children") can never have a combined deviance that is greater than the deviance of the node they were created from ("parent"). The split is made by considering all of the possible divisions of the variables used in constructing the tree, and choosing the split that maximizes the decrease in deviance. The procedure is continued recursively until the size of a "child" node would be forced below a preset threshold or deviance can not be decreased by a preset threshold. The final product is a tree with a number of terminal points (called leaves) that predict the success or failure of the individuals in the terminal nodes based on the response that holds the majority in that node. The final deviance is the sum of the deviances for all the terminal nodes (Venables and Ripley, 1994).

Once a tree is constructed, it typically contains a large number of nodes, making it too closely fitted to the data and therefore not accurate for use in prediction. To correct this, there are procedures that can be used to reduce the tree to an optimal size.

The most common procedure is the use of cross-validation. This procedure randomly splits the data into ten equally-sized sets. Nine of the sets are used to grow a tree and the tenth is used to test the tree and determine deviance. This is completed on all ten possible permutations of the divided data. The tree size with the lowest average deviance is considered the best size tree. It must be noted that since the data is divided randomly, the results may differ slightly for separate cross-validations on the same data (Venables and Ripley, 1994). This optimum-sized tree is the final result of growing a classification tree and can be used for prediction.

III. DATA COMPARISONS

At the start of the data comparisons, the data set consisted of 86,815 records. These are for the recruits that had entered the Navy in FY's 95-96 at least two years prior to the compiling of the data set. This data set was then checked for possible data errors. Three possible data errors were found in the fields pertinent to this chapter.

First, there is an error in the Navy Loss Codes. Loss code 833 is listed in two separate places. In one location it was listed as a non-attrition loss identified as "Honorable discharge, Unsuitability-Homosexual." In the other location it was listed as an attrition discharge identified as "Honorable discharge, misconduct" which is also identified as unsuitability attrition for this study. It was determined to classify all losses with code 833 as unsuitable attrition. Since the non-attrition discharge of people with homosexual preferences is for a violation of Navy policy, it was decided that making this discharge "unsuitable" was the proper way to account for the code being listed in two separate loss classifications.

The second error was in the attrite codes. It was found that the number of recruits identified with a 1 (attrited in two years) in the attrite code were not the same as the number of recruits with fewer than 730 days of active service as identified by the service days code. Since more enlistees served fewer than 730 days then were listed as attrites, the initial belief was that the former were two-year enlistees who left the Navy 90 days early. However, when this was checked it was not the case. When looking at the recruits that were not identified as attrites but had fewer than 730 days of service, it was found

that they were spread throughout all enlistment programs. Due to this, it was determined to use service day as the determination of attrition, with a recruit that had fewer than 730 days of active service classified as a two-year attrite.

The third error was in the accession waiver category. This section had two errors in it. The first was in the classification of moral waivers as "other" and "N/A" in the second character of codes that begin with "D"(moral waiver). Although the codes for types of moral waivers appear to be well-defined, there are recruits identified with the other and N/A codes. It was determined to treat these as separate waiver categories. The second error in the moral waiver codes is the identification of a waiver for fewer than three minor misdemeanors, when guidance was that a waiver is required only for three or more minor misdemeanors. The first check on this error was to determine if this was because of the stricter requirements for program waivers. However, a sample of recruits in this category did not have all members belonging to rates or programs that required program waivers. Therefore, it was determined to treat this as a separate category of waiver type. It is the belief that these errors result from the fact that waivers are requested at the recruiter level and were either misreported or requested when not needed.

These were the only data errors found that affect the comparisons of attrition for recruits with moral waivers against those without. None of these errors was found to be serious enough to prevent the use of the data for the study. Therefore, attrition comparisons were conducted for both the entire data set and the modified data set.

A. ENTIRE DATA SET

The entire data set consists of 86,815 records. To start, recruits with moral waivers and recruits without moral waivers were separated into two separate data sets. It was found that 12,464 of the recruits had moral waivers. Table 5 provides the breakdown of moral waivers by waiver category.

Table 5: Entire Data Set Waiver Breakdown

Type	Number	% of data set	% of moral waivers
Total	86815	100%	-----------------------
No moral waiver	74351	85.64%	-----------------------
Moral waivers	12464	14.36%	100%
- minor traffic	44	0.05%	0.35%
-<3 minor misdemeanors	289	0.33%	2.32%
->=3 minor misdemeanors	122	0.14%	0.98%
-non-minor misdemeanors	7858	9.05%	63.05%
-felony(adult)	42	0.05%	0.37%
-felony(juvenile)	48	0.06%	0.39%
-drug related	3403	3.92%	27.30%
-alcohol related	564	0.65%	4.52%
-other	66	0.08%	0.53%
-N/A	28	0.03%	0.22%

The first noticeable point in Table 5 is that the percentage of moral waivers, 14.36%, is much lower than anticipated by previous studies. It is also found that 63% of the moral waivers are for non-minor misdemeanors and when non-minor misdemeanors and drug-related waivers are combined, they make up over 90% of the moral waivers. The unexplained categories of other and N/A make up less than 1% of the moral waivers; therefore their misclassification should be insignificant to the overall study.

It is also important to determine the attrition percentages for the moral and non-moral waiver recruits. Table 6 provides the losses per Navy loss code, grouped into non-attrition losses, attrition but not unsuitable losses, and unsuitable attrition losses for both moral waiver recruits and for those without moral waivers. Codes with zero losses are excluded from the table.

Table 6: Entire Data Set Loss Code Breakdown
(Non-Attrition and Attrition not Unsuitable)

	No Moral Waiver		Moral Waiver	
Type	Number	% of total recruits	Number	% of total recruits
TOTAL LOSSES	23522	31.64 %	4996	40.08%
Non attrition losses	514	0.69 %	80	0.64%
801	2	0.003%	1	0.01%
816	1	0.001%	0	0.00%
931	1	0.001%	0	0.00%
942	235	0.32 %	23	0.18%
943	162	0.22 %	24	0.19%
980	112	0.15 %	32	0.26%
998	1	0.001%	0	0.00%
Attrition not unsuitable	4641	6.24 %	676	5.42%
804	498	0.67 %	60	0.48%
805	420	0.56 %	68	0.55%
808	2	0.003%	0	0.00%
813	2051	2.76 %	289	2.32%
814	136	0.18 %	21	0.17%
830	2	0.003%	0	0.00%
844	12	0.02 %	1	0.01%
845	14	0.02 %	1	0.01%
853	816	1.10 %	144	1.16%
854	10	0.01 %	0	0.00%
933	152	0.20 %	28	0.22%
951	171	0.23 %	21	0.17%
952	79	0.11 %	14	0.11%
954	1	0.001%	0	0.00%
958	60	0.08 %	3	0.02%
959	217	0.29 %	26	0.21%

Table 6 (Contd.): Entire Data Set Loss Code Breakdown

(Unsuitable Attrition)

	No Moral Waiver		Moral Waiver	
Type	Number	% of total recruits	Number	% of total recruits
Unsuitable attrition	18367	24.70 %	4240	34.02%
817	2	0.003%	0	0.00%
818	131	0.18 %	54	0.43%
831	189	0.25 %	55	0.44%
833	184	0.25 %	30	0.24%
857	12	0.02 %	4	0.03%
858	221	0.30 %	56	0.45%
870	12	0.02 %	5	0.04%
871	1449	1.95 %	359	2.88%
872	62	0.08 %	7	0.05%
887	580	0.78 %	126	1.01%
888	3678	4.95 %	1037	8.32%
890	22	0.03 %	3	0.02%
901	58	0.08 %	17	0.14%
902	1	0.001%	0	0.00%
970	11046	14.86 %	2227	17.87%
971	719	0.97 %	260	2.09%
972	1	0.001%	0	0.00%

The first analysis of Table 6 is to look at the effect of the identified data errors. Code 833 that had been identified as a non-attrition loss and an attrition loss, but converted to an unsuitable attrition loss, does not appear to have a significant effect on the study. Among recruits with this code, the difference in attrition rates between recruits with moral waivers and those without is 0.01 percentage points.

The differences in rates between recruits with moral waivers and those without, shown in Table 6, were tested for significance by comparing the population proportions that were found for each set. The comparison is conducted to see if the difference seen is large enough to be called significantly different or if it is possibly just a result of chance.

(Here the analysis proceeds as if each group was a random sample from some "super-population" of potential recruits; the interest is in whether these two populations have different attrition rates.) Tests for significance were conducted using $\alpha = 0.01$. The overall loss rate was found to be significant (using z-test for proportions, $p = 0.0000$) at 8.4 percentage points higher for recruits with moral waivers. It is also found that non-attrition losses are not significantly different (using z-test for proportions, $p = 0.2654$) between groups. The rates for "attrition but not unsuitable" losses are significantly different (using z-test for proportions, $p = 0.0002$) but the percentage difference is less than a full percentage point. The recruits without moral waivers have a higher loss rate in this category than do the recruits with waivers. However, when the codes are looked at individually, only one (Code 813) has a significant difference (using z-test for proportions, $p = 0.0025$) of the codes that are large enough to use the normal approximation. The cause of this difference is unknown, with recruits without moral waivers having the higher loss percentage.

The unsuitable attrition losses do show a significant difference (using z-test for proportions, $p = 0.0000$). Recruits with moral waivers have a 9.3 percent point (34.02%) higher attrition rate than the recruits without moral waivers (24.70%). It is also noted that the majority of unsuitable attrition for both groups is entry level separations (Code 970), and recruits with moral waivers have a 3 percentage point higher attrition in this category. Undesirable discharge-misconduct (Code 888) is the second largest in each group, but the size of this group is less than half the size of the larger Code 970 discharges. Each of

these two codes also shows significance (using z-test for proportions, $p = 0.0000$ for each) when moral waiver recruits and recruits without moral waivers are compared.

Since there does appear to be significantly higher unsuitable attrition for recruits with moral waivers, comparisons are also conducted for each type of waiver. Table 7 provides the percentages of losses for each of the loss categories separated by waiver category.

Table 7: Entire Data Set Waiver Category Breakdown

Waiver category	Total Number	Total Loss %	Non-Attrition Loss %	Attrition Not Unsuitable %	Unsuitable %
Minor traffic	44	36.36%	0.00%	4.55%	31.82%
<3 minor misdemeanors	289	32.18%	0.35%	7.61%	24.22%
>=3 minor misdemeanors	122	33.61%	0.00%	5.74%	27.87%
Non-minor misdemeanors	7858	41.40%	0.75%	4.90%	35.75%
Felony(adult)	42	42.86%	0.00%	4.76%	38.10%
Felony(juvenile)	48	41.67%	0.00%	4.17%	37.50%
Drug-related	3403	38.73%	0.53%	6.85%	31.35%
Alcohol-related	564	35.46%	0.00%	3.37%	32.09%
Other	66	45.45%	1.52%	6.06%	37.88%
N/A	28	25.00%	3.57%	0.00%	21.43%

When analyzing the individual waiver categories, the non-attrition losses and "attrition but not unsuitable" losses are similar to results of the entire data. The other and N/A categories are different, but these categories have few data points so the expected variability is high.

The unsuitable attrition percentages do provide insight on the higher attrition rates of the moral waiver recruits. Recruits with fewer than 3 minor misdemeanors (24.22%) and the N/A group (21.43%) have unsuitable attrition rates that are similar to the no-

waiver recruits (24.70%). This is important since these groups do not appear to actually require waivers, but the "other" category that doesn't require waivers does not have low rates and has one of the highest unsuitable attrition rates (37.88%). The 3 or more minor misdemeanors unsuitable attrition rate (27.87%) falls between the unsuitable attrition rates of the waiver (34.02%) and non-waiver groups (24.70%), with all other waiver categories near or above the rate of the waiver group. The two felony categories and the other category are higher than the overall moral waiver group, but these three groups also have very small sample sizes.

When looking at the unsuitable attrition percentages, it is apparent that they are higher for the moral waiver group. All of the categories that were identified as requiring waivers contribute to the higher rates, with two of the three waivers that do not appear to be required having rates that are near the no moral waiver rate.

B. MODIFIED DATA SET

For this section, the data set of 86,815 records was modified by removing the rates and programs identified in Table 4. This was done to remove the influence of program waivers from the data. The resulting database contained 56,510 records that were then separated into moral waiver and non-moral waiver data sets. It was found that 7,767 of the recruits had moral waivers in this data set. Table 8 on the next page summarizes the breakdown of moral waivers and lack of moral waivers for this data set.

When analyzing Table 8, it is again apparent that the percentage of moral waivers is lower than expected at 13.74%. Non-minor misdemeanors are the dominant moral

waivers, comprising 77.83% of the moral waivers in this database. Again it is noted that the unexplained categories of other and N/A make up less than 1% of the moral waivers in this group. Based on this, the vagueness of these categories should be insignificant to the study.

Table 8: Modified Data Set Waiver Breakdown

Type	Number	% of data set	% of moral waivers
Total	56510	100%	----------------------
No moral waiver	48743	86.26%	----------------------
Moral waivers	7767	13.74%	100%
- minor traffic	19	0.03%	0.24%
-<3 minor misdemeanors	74	0.13%	0.95%
->=3 minor misdemeanors	103	0.18%	1.33%
-non-minor misdemeanors	6045	10.70%	77.83%
-felony(adult)	35	0.06%	0.45%
-felony(juvenile)	36	0.06%	0.46%
-drug related	1074	1.90%	13.83%
-alcohol related	328	0.58%	4.22%
-other	48	0.08%	0.62%
-N/A	5	0.009%	0.06%

The next step is to determine attrition percentages for all recruits, broken into non-waiver and waiver categories. Table 9 provides losses per Navy Loss Code, grouped by non-attrition losses, attrition but not unsuitable losses, and unsuitable attrition losses. Codes with zero losses are excluded from the table.

The first point to consider in Table 9 is the effect of the data error in Code 833. The placement of this code does not significantly affect this data set, since the attrition is nearly the same for both the waiver and non-waiver groups and the attrition for this code is less than 0.25% for both groups.

Table 9: Modified Data Set Loss Code Breakdown

(Attrition not Unsuitable and Unsuitable Attrition)

	No Moral Waiver		Moral Waiver	
Type	Number	% of total recruits	Number	% of total recruits
TOTAL LOSSES	15949	32.72%	3314	42.67%
Attrition not unsuitable	2722	5.58 %	353	4.54%
804	309	0.63 %	32	0.41%
805	257	0.53 %	38	0.49%
808	1	0.002%	0	0.00%
813	1270	2.61 %	157	2.02%
814	96	0.20 %	13	0.17%
830	1	0.002%	0	0.00%
844	9	0.02 %	0	0.00%
845	9	0.02 %	1	0.01%
853	454	0.93 %	68	0.88%
854	7	0.01 %	0	0.00%
933	93	0.19 %	15	0.19%
951	138	0.28 %	12	0.15%
952	49	0.10 %	12	0.15%
958	3	0.006%	0	0.00%
959	26	0.05 %	5	0.06%
Unsuitable attrition	12838	26.34 %	2894	37.26%
817	2	0.004%	0	0.00%
818	73	0.15 %	31	0.40%
831	144	0.30 %	36	0.46%
833	114	0.23 %	19	0.24%
857	11	0.02 %	4	0.05%
858	123	0.25 %	34	0.44%
870	7	0.01 %	3	0.04%
871	1024	2.10 %	248	3.19%
872	36	0.07 %	2	0.03%
887	347	0.71 %	83	1.07%
888	2646	5.43 %	710	9.14%
890	14	0.03 %	0	0.00%
901	43	0.09 %	14	0.18%
902	1	0.002%	0	0.00%
970	7718	15.83 %	1521	19.58%
971	534	1.10 %	189	2.43%
972	1	0.002%	0	0.00%

Table 9 (Cont.): Modified Data Set Loss Code Breakdown

(Non-Attrition losses)

	No Moral Waiver		Moral Waiver	
Type	Number	% of total recruits	Number	% of total recruits
Non attrition Losses	389	0.80 %	67	0.86%
801	2	0.004%	1	0.01%
816	1	0.002%	0	0.00%
931	1	0.002%	0	0.00%
942	136	0.28 %	17	0.22%
943	159	0.33 %	24	0.31%
980	90	0.18 %	25	0.32%

The overall difference in loss percentages between the moral waiver and non-moral waiver groups is significant (using z-test for proportions, $p = 0.0000$) in this data set. Significance was again tested using $\alpha = 0.01$. The overall loss rate for recruits with moral waivers (42.67%) is 9.9 percentage points higher than that of the recruits without moral waivers (32.72%). However, the loss percentages for non-attrition losses are not significantly different (using z-test for proportions, $p = 0.2915$) between the groups. Also, the loss percentages for the "attrition but not unsuitable" losses do show a significant difference (using z-test for proportions, $p = 0.0001$) as it is 1 percentage point higher for the non-waiver group. However, only one code (Code 813) has a difference that is considered significant (using z-test for proportions, $p = 0.0010$) when the codes that are large enough for the normal approximation are looked at individually. This is the same code that shows significance in the entire data set and the significance is therefore caused by the same unknown reasons, since this data set is a subset of the first data set.

The unsuitable attrition rate for the moral waiver group is 10.9 percent (37.26%) higher than that of the non-moral waiver recruits (26.34%), which is a statistically significant difference (using z-test for proportions, $p = 0.0000$). Within the unsuitable attrition category, the majority of the individual loss codes are higher for the moral waiver recruits. However, five of the codes do have larger attrition rates for the non-waiver recruits, the largest difference among them being 0.04 percentage points. In contrast, there are four codes for which the moral waiver group has an attrition rate that is greater by more than 1 percentage point. Entry level separation (Code 970) and undesirable misconduct discharge (Code 888) have the largest differences with the moral waiver group having an unsuitable attrition rate that is more than 3.7 percentage point higher than the non-waiver group for each. Codes 970 and 888 also have the largest attrition numbers in each group, with entry level separations being the dominant reason for unsuitable attrition in both groups.

Since Table 9 shows a significant difference in loss percentages and unsuitable attrition between recruits with moral waivers and those without, each waiver category will also be analyzed. Table 10 provides the percentage of losses for each category, broken down by waiver types.

Within the individual waiver categories, it is first noticed that the non-attrition losses occur for only two types of waivers. It is found that none of the categories have a significant difference, with $\alpha = 0.01$, from the non-attrition losses of recruits without waivers. With respect to the attrition but not unsuitable losses, the percentage loss is sizably lower than the overall no waiver group in Table 9 (5.58%) for the alcohol-related

Table 10: Modified Data Set Waiver Category Breakdown

Waiver category	Total Number	Total Loss %	Non-Attrition Loss %	Attrition Not Unsuitable %	Unsuitable %
Minor traffic	19	47.37%	0.00%	5.26%	42.11%
<3 minor misdemeanors	74	41.89%	0.00%	5.41%	36.49%
>=3 minor misdemeanors	103	32.04%	0.00%	5.83%	26.21%
Non-minor misdemeanors	6045	42.60%	0.93%	4.52%	37.15%
Felony(adult)	35	48.57%	0.00%	5.71%	42.86%
Felony(juvenile)	36	41.67%	0.00%	5.56%	36.11%
Drug related	1074	45.07%	1.02%	5.03%	39.01%
Alcohol related	328	38.72%	0.00%	2.44%	36.28%
Other	48	45.83%	0.00%	6.25%	39.58%
N/A	5	20.00%	0.00%	0.00%	20.00%

waivers (2.44%) and the N/A category has a rate of 0.0% due to a small sample size. No category has a rate that stands out. It is noted that none of the rates tests as significantly different from the attrition but not unsuitable loss percentage of the non-waiver group as a whole. The relatively small variation of the categories from the rate of the non-waiver group appears to support the belief that there is no difference between the attrition but not unsuitable loss percentages of the two groups.

When analyzing the separate waiver categories of unsuitable attrition, all but two are higher than the unsuitable attrition of the non-waiver group (26.34%). N/A is lower (20.00%), but with only five recruits in the category, it is not tested for significance nor considered critical. Also, the rate for recruits with 3 or more minor misdemeanors (26.21%) is very near the rate of the recruits without moral waivers. Only three of the waiver categories (Non-minor misdemeanors, drug and alcohol related) have unsuitable

attrition rates that are significantly higher (using z-test for proportions; p = 0.0000 for all three) than the rates for recruits without moral waivers. Minor traffic and adult felonies have the highest rates, with unsuitable attrition rates above 42% and overall losses above 47%. Overall, this data set again supports the belief that unsuitable attrition is significantly higher for recruits with moral waivers, with 8 of the 10 categories showing this higher loss.

C. DATA SET COMPARISONS

In comparing the data sets, it must first be noted that the modified data set is a subset of the other data set. There are three important changes seen in the modified data set. First, the percentage of recruits receiving moral waivers decreases slightly. This is expected since this data set was created by removing recruits with program waivers.

The second change is an adjustment in the types of waivers granted. The percentage of recruits receiving drug-related waivers was cut in half. This left non-minor misdemeanors as 77% of the moral waivers, compared to 63% in the entire data set. This change is again thought to be a result of the removal of program waivers, since the majority of program waivers are for drug-related offenses due to security clearance issues.

The third change is that the overall loss percentages and unsuitable attrition percentages are higher for all subsets of recruits in the modified data set. However, the difference between moral and non-moral waiver recruits is larger in the modified data set than it is in the entire data set. For example, the group with moral waivers has an unsuitable attrition rate that is 10.9% higher than the non-waiver group in the modified

data set and only 9.3% higher in the original data set. This leads to the assumption that program waivers are causing the difference in unsuitable attrition rates to appear smaller than it really is.

IV. PREDICTION MODELS

The records of recruits with moral waivers were used to develop models to predict success or failure of future recruits with a moral waiver. The first step in this process was to modify the data sets for use in the prediction models. The modifications were conducted on both the entire data set and the modified data set. Therefore they will be explained prior to discussing the individual data sets.

The first modification was to construct a field for 2-year unsuitability attrition. In making this field, the first step was to identify recruits with fewer than 730 service days. These recruits were then checked for a loss code that corresponded to unsuitable attrition. Recruits with less than 730 service days and an unsuitable attrition loss code were identified with a "yes" and all other recruits with a "no."

The next modifications were to the race/ethnic codes and the recruiting districts. First, race and ethnic codes were combined to create a Hispanic ethnicity within the race codes. To do this modification, recruits with ethnicity codes of 1, 4, 6, 9, and S (code definitions in Appendix B) had their identifiers in the race code changed to Hispanic. All other recruits kept their previous race codes. Recruiting regions were also created by grouping the recruiting districts into their respective regions (per Appendix B).

Another new data field was age at entry into the Navy. This was computed by taking the date of accession into the Navy (CANDATE) and subtracting the date of birth (DOB) for each recruit. The result was then divided by 365.25 and rounded down to compute the age at entry.

The next step of data set modification was creating the moral waiver categories. These were made by using the second letter of the waiver categories, which corresponded to the type of the moral waiver granted. Because of small sizes and close relationships between categories, some waiver categories were combined. "Other" (category X) and "N/A" (category Y) were combined into a single category. The two felony categories (categories E and F) were also combined into a single waiver category.

One final modification was conducted prior to checking for data errors. The PAYGRADE field was modified so that all recruits who entered the Navy as E-3 or greater were put into the E-3 category. This left 13 variables and the response for use in the predictive models. The variables are DEPDAYS, AFQT, TERM, EDCERT, SEX, DEPEND, Recruiting Region, PRIOR_SV, New Race/Ethnic code, age at entry, Paygrade, PROGRAM, and Moral Waiver code. The response is the 2-year unsuitability attrition code that was constructed.

The data was then checked for errors. Age at entry was the first field to be checked for errors. Records were found that had ages above 60 years old and with N/A age entries. It was found that these errors came from birth date entries that appeared to be in error or were blank. Records without birth dates or that had ages greater than 60 were removed. Upon completion of all data corrections and modifications, there were no ages above 34.

The next set of errors was in the DEPDAYS field. It was found that some of the records had N/A entries, which had resulted from no DEP entry date for the recruit. It was assumed that no DEP entry date was the result of the recruits never entering DEP.

Based on this assertion, all of the N/A DEPDAYS were converted to 0. This was accomplished by writing an S-Plus function to do the conversion. This is a modification to an existing S-Plus function (na.gam.replace) and is included in Appendix D.

Another data error was found in the AFQT scores. Data entries were found with test scores of zero. However, further analysis showed that all of the recruits with zero test scores were recruits with prior military service. Over half of the prior service recruits in each data set had test scores of zero. When analyzing the prior service recruits, it was found that none of them attrited due to unsuitable attrition as defined for this study. Therefore, it is easy to determine that prior service is a very positive attribute when recruiting recruits with moral waivers. It is also determined that since this is so easily determined, prior service recruits can be removed from the data set to correct the AFQT score error.

One other issue was found in initial modeling. The variables TERM and PROGRAM are related to each other. For each PROGRAM, there is a specific TERM associated to it. Therefore, they can not both be used in the logistic model. TERM was removed from the variables to be used in the prediction models.

No other errors were found in the data to be used for prediction models. This results in having eleven predictive variables and the response for use in the models. The remaining eleven variables to be used are DEPDAYS, AFQT, EDCERT, SEX, DEPEND, Recruiting Region, New Race/Ethnic code, age at entry, new Paygrade, PROGRAM, and Moral Waiver Code.

A. ENTIRE DATA SET

The original entire data set consisted of the records for 12,464 recruits who received moral waivers. From this set, 4 records were removed due to age errors and 167 prior service recruits records were removed. This leaves a data set of 12,293 records for use in the predictive models. Of these records, 320 had N/A values in the DEPDAYS column which were converted to 0. There were also two recruits that entered at pay grades above E-3 who were grouped with the E-3 entrants for this study. These changes result in the final data set to be used in the prediction models for the entire data set.

Another important aspect of setting up the data set is variable classifications. Variables in this data set are of two types: Numeric and Factor. Numeric variables are variables that take continuous numbered values over a range, whereas a factor is a categorical variable that takes specific values or names. Table 11 provides the classification of the variables in this model and the number of factors when applicable. The number of factors is the actual number of categories for the variable. However, when

Table 11: Entire Data Set Prediction Variables

Variable	Type	Number of Levels
AFQT	Numeric	N/A
EDCERT	Factor	3
SEX	Factor	2
PROGRAM	Factor	11
Recruiting Region	Factor	4
Race/Ethnic	Factor	6
PAYGRADE	Factor	3
Age at entry	Numeric	N/A
DEPDAYS	Numeric	N/A
DEPEND	Numeric	N/A
Waiver type	Factor	8

modeled in the logistic regression, one of the factors for each variable will be set as the baseline. All of the other factors will be modeled in relation to the baseline factor.

1. Logistic Model

The logistic model was started by modeling Unsuitable Attrition using all of the predictive variables. The resulting model was then subjected to an analysis of variance (ANOVA) using a χ^2 (chi-squared) test for each variable (Hamilton, 1992). The goal was to create a model using the variables that decrease deviance the most. The χ^2 test uses the change in deviance and degrees of freedom to test the significance of the variable, where the deviance is an indicator of the variance associated with the variable of interest. The test of significance in the ANOVA is a test of the hypothesis that the coefficients associated with a variable are equal to 0. If the test fails to accept the hypothesis that the coefficients are equal to 0, then the variable is used in the model. The determination of the acceptance of a variable is conducted by comparing the *p*-value of the variable with a pre-determined acceptance value (α).

The ANOVA table was used to reconstruct the model in an ascending order of χ^2 *p*-values. A new model was then created and ANOVA run on it to test *p*-values. Variables with *p*-values above 0.05 using the χ^2 test were removed and new models developed until no variables remained with an ANOVA χ^2 *p*-value above 0.05. Table 12 shows the final ANOVA *p*-values for the variables remaining in the model. Table 13 provides the prediction values for the final logistic model.

Table 12: Entire Data Set Final ANOVA

Variable	P-value (χ^2)
AFQT	0.0000
EDCERT	0.0000
SEX	0.0000
PROGRAM	0.0000
Recruiting Region	0.0000
Race/Ethnic	0.0000

Table 13: Entire Data Logistic Prediction Coefficients

Variable	Coefficient	Variable	Coefficient
Intercept	-0.425195	EDCERT: D	Baseline
AFQT	-0.010841	G	0.612859
SEX: F	Baseline	N	0.727540
M	0.438342	PROGRAM: 2YO	Baseline
Region: East	Baseline	3YO	0.123553
North	0.021932	5YO	-0.128990
South	-0.077030	AEF	0.157441
West	-0.218449	ATF	0.048042
Race/Ethnic: C	Baseline	DIVR	-0.003982
H	-0.119489	JOBS	0.379989
M	-0.857296	NF	-0.164071
N	-0.011231	SF	0.370783
R	0.277597	SG	0.133833
X	0.004019	TEP	-0.007066

When looking at the prediction coefficients, the sign of the result is the most critical point. A positive result means that the predictor increases the chance of unsuitable attrition and a negative result indicates the opposite. For variables that are factors, the baseline level is set at 0 and the other levels of that variable are in relation to the baseline level.

The most noteworthy result of the logistic model is the effect of EDCERT on the probability of attrition. In EDCERT, D (Diploma grad) is set as the baseline and both G (G.E.D.) and N (non-grad) have a very high increase in the chance of unsuitable attrition when compared to diploma grads. The difference can also be seen in an example that shows how to compute the probability of unsuitable attrition. This is computed by setting all of the variables but EDCERT to a given value and varying EDCERT between D (Diploma grad) and N (non-grad). The variable values for the example are AFQT of 70, SEX of Female, North Recruiting Region, Hispanic (H) Race/Ethnic code, and 3YO PROGRAM. With these, the attrition probability can be computed. For recruits with EDCERT of N (non-grad):

$$\hat{p} = 1/(1+\exp(-(-0.425195+(70*-0.010841)+0+0.021932-0.119489+0.123553+0.727540)))$$
$$\hat{p} = .39$$

For recruits with EDCERT of D (diploma grad) it is:

$$\hat{p} = 1/(1+\exp(-(-0.425195+(70*-0.010841)+0+0.021932-0.119489+0.123553+0)))$$
$$\hat{p} = .24$$

This results in a difference of 0.15 in attrition percentage probabilities, showing that, according to the model, recruits with high school diplomas are more likely to succeed than the recruits from the non-grad group, since the non-grad recruit has a higher probability of unsuitable attrition.

There are two other important points found in the results. One is in the Race/Ethnic factors. Code M (Asian) in the Race/Ethnic factors is associated with a significant decrease in the chance of unsuitable attrition. It is also noted that the SEX

factor of M (male) is associated with a notable increase in the chance of unsuitable attrition.

The biggest use of the logistic model lies in its use as a predictive tool. To determine its success as a prediction tool, we must compare it to the "naïve model," which is the predicted unsuitable attrition rate when there is no model. The naïve model attrition percentage is 34.49%. It is important to note that this is different than the unsuitable loss percentage in Chapter III because of the removed data.

There are two possible ways to determine error for the logistic model. First, we must remember that the model returns a probability between 0 and 1. Therefore, one option is to use the midpoint of 0.5 as the point that determines if a recruit is predicted to attrite (That is, a predicted probability greater than 0.5 results in a prediction of attrition, and less than 0.5 results in a prediction of no attrition.). The other is to pick a break point that minimizes the error and use that point to determine whether the recruit is predicted to attrite. An S-plus function that was written to determine the point of minimum error is included in Appendix D. The error for both cases is the sum of the recruits we predicted as completing two years who attrited and recruits we predicted would attrite who did not divided by the number of recruits in the data set. The recruits that are predicted to survive are the recruits with a value below 0.5 or below the calculated breakpoint. For the 0.5 model, the error is 34.36% and the minimum error is 34.24% at the prediction break point of 0.54. Neither of these results shows a major improvement over the naïve model.

Another important aspect of the model is to look at the number of recruits that would not have been accepted for enlistment based on the model. For the 0.5 model, 592

recruits would not have been accepted. Of the 592, 288 did not attrite and 304 did, a 48.65% error caused by not enlisting the 288 recruits who succeeded. For the 0.54 model, 227 recruits would not have been enlisted, 98 that did not attrite and 129 who did, a 43.17% error.

The last step is to validate the model. This model was built using all of the available data to ensure all factors were incorporated into the model. Therefore, the data used to build the model is the only data available to test the model. A modified cross-validation procedure is used to test the model. This procedure is in an S-Plus function, included in Appendix D. This procedure takes the data and divides it into 10 subsets. Models are developed with each subset omitted and misclassification errors when the model is applied to the held-out data are accumulated. The function then returns the accumulated average misclassification error. The goal is for this error to be near the error found by the model on the entire data to ensure that the model was not over-fitted to the data. This error is 34.56% for the 0.54 model and 34.36% for the 0.5 model. These do not vary largely from the prediction found on the entire data; therefore the model is determined to be acceptable.

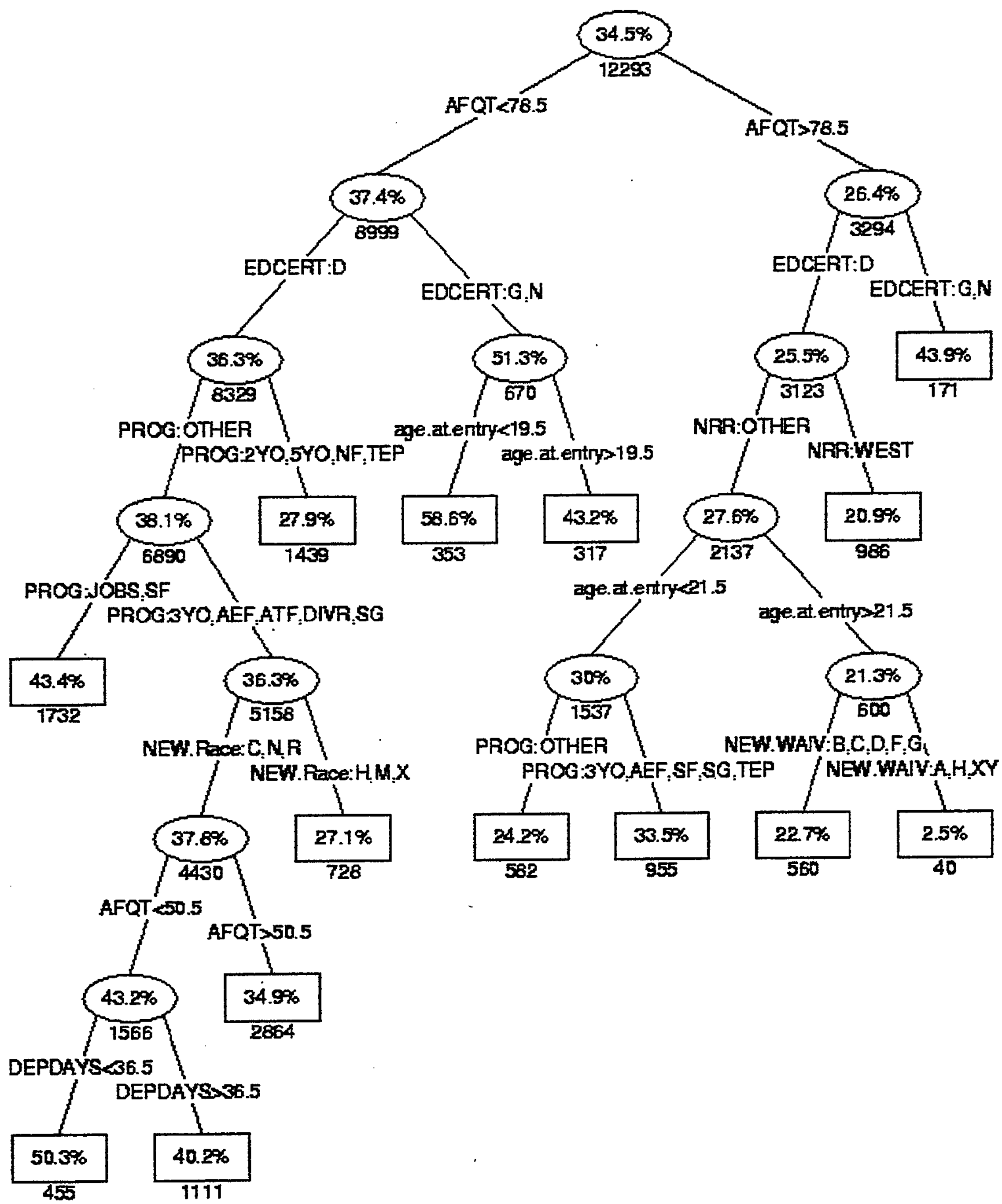
2. Classification Tree

The classification tree was started by modeling Unsuitable Attrition by the prediction variables in Table 11. The tree was constructed as discussed in Chapter II, and then had a cross-validation conducted on it to find the best-sized tree. The cross-validation was run until the size of the tree with the minimum deviance was found twice, since the size can vary based on the split the model chooses. For this model, the size is 14

and is found in two cross-validation runs. The tree is then pruned to size 14 and used as the classification tree model. Figure 3 is the graphical result, where the number below each node shows the number of recruits in that node. The number in the node is the percentage of recruits in that node who are from the "unsuitable attrition" group (yes's). For example, of those recruits whose AFQT score was smaller than 78.5, 37.4% of the 8999 recruits with these scores underwent unsuitable attrition.

This tree has a 33.97% misclassification error and there are 808 recruits who would not have been enlisted. Of those 808 recruits, 372 (46.04%) did not attrite and 436 did. The recruits that would not have been accepted are from two groups. The first group includes recruits with AFQT score of less than 50.5, an EDCERT code of D (diploma grad), in an enlistment program of 3YO, AEF, ATF, DIVR, or SG, a Race/Ethnic code of Caucasian (C), Black (N), or American Indian (R), and with DEPDAYS less than 36.5 days. The second group includes recruits with AFQT score of less than 78.5, an EDCERT code of G.E.D. (G) or non-grad (N), and an age at entry of less than 19.5. Since cross-validation was used to build this model, the model's predictive ability was validated during its construction.

Figure 3: Entire Data Set Classification Tree



3. Model Results

Neither the classification tree nor the logistic regression provides an improvement of any size over the naïve model. The classification tree has the best improvement in error rate, but also excludes the most recruits. The other important feature of these two models is that the EDCERT variable shows an importance in both the models.

B. MODIFIED DATA SET

This data set started with the records of the 7,767 recruits in the modified data set who received moral waivers. This data set is a subset of the entire data set. In this data set, 3 records were removed due to age errors and 91 records of prior service recruits were removed. This left a data set of 7,673 records for use in the prediction models. In this data set, 220 records had N/A values in DEPDAYS, which were converted to 0. There was also two recruits that entered at pay-grades above E-3, who were grouped with the E-3 recruits. This left the final data set to be used for the modified data set models.

The other important aspect of this model is variable classifications. The variables are classified in the same way as for the entire data set model. Table 11 from the first model provides the classification of all the variables. The number of factors also matches Table 11 in all but one case. The special case is the PROGRAM variable that is modified to 6 factors for this model.

1. Logistic Model

The modified data logistic model was developed using the same procedure as was used for the entire data set. This model was created using iterative steps of building the

model and conducting ANOVA χ^2 tests. Once all of the variables with *p*-values greater than 0.05 were removed, the final logistic model was complete. Table 14 provides the final ANOVA and *p*-values. Table 15 provides the prediction values for the logistic model.

Table 14: Modified Data Set Final ANOVA

Variable	P-value (χ^2)
AFQT	0.0000
EDCERT	0.0000
SEX	0.0000
PROGRAM	0.0000
Recruiting Region	0.0000
Race/Ethnic	0.0002
Waiver type	0.0322

Table 15: Modified Data Logistic Prediction Coefficients

Variable	Coefficient	Variable	Coefficient
Intercept	-0.311746	AFQT	-0.008116
EDCERT: D	Baseline	PROGRAM: 2YO	Baseline
G	0.442494	3YO	0.151545
N	0.688349	5YO	-0.099184
Region: East	Baseline	SF	0.383044
North	-0.024071	SG	0.165421
South	-0.109614	TEP	0.017647
West	-0.260274	Waiver: A	Baseline
Race/Ethnic: C	Baseline	B	-0.330960
H	-0.142049	C	-0.823980
M	-0.785563	D	-0.275740
N	-0.011494	EF	-0.196391
R	0.335328	G	-0.075579
X	0.178342	H	-0.296479
SEX: F	Baseline	XY	-0.155761
M	0.468322		

In these results, it is important to remember the impact of the coefficient's sign. A positive sign indicates that the predictor increases the chance of unsuitable attrition and a negative sign decreases the chance. For variables that are modeled as factors, the results are in relation to the baseline characteristic.

When looking at these results there are some important points that are visible. The first is in the EDCERT variable, where it is evident that codes N (non-grads) and G (G.E.D.) have a greater chance of unsuitability attrition then high school grads, with non-grads having the higher increase. It is also noted that race code M (Asian) strongly decreases the probability of unsuitable attrition. The probability of unsuitable attrition is higher for men then for women. The other important aspect of the prediction coefficients is in the waiver types. First, all of the coefficients are negative, meaning the baseline (category A-minor traffic) increases the chance of unsuitable attrition among the waiver types. The other interesting fact is that category C (3 or more misdemeanors) has a very large impact on decreasing the chance of unsuitable attrition when compared to the other categories.

The importance of the results can also be shown in an example. The example will compute the changes in probability of unsuitable attrition caused by a change in the EDCERT variable. This is computed by setting all of the variables but EDCERT to a given value and varying EDCERT between D (Diploma grad) and N (non-grad). The variable values for the example are AFQT of 70, SEX of Female, North Recruiting Region, Hispanic (H) Race/Ethnic code, waiver category A, and 3YO PROGRAM. With

these, the attrition probability can be computed. For recruits with an EDCERT code of N (non-grad):

$$\hat{p} = 1/(1+\exp(-(-0.311746+(70*-0.008116)+0-0.024071-0.142049+0+0.151545+0.688349)))$$
$$\hat{p} = .45$$

For recruits with EDCERT of D (diploma grad) it is:

$$\hat{p} = 1/(1+\exp(-(-0.311746+(70*-0.008116)+0-0.024071-0.142049+0+0.151545+0)))$$
$$\hat{p} = .29$$

This results in a difference of 0.16 in attrition percentage probabilities, showing that recruits with high school diplomas are more likely to succeed than the recruits from the non-grad group, all other things being equal.

However, the most important aspect of the logistic model is its prediction capability. To determine this, it is compared to the "naïve model", which is the unsuitable attrition when there is not a model. The unsuitable attrition for the naïve model is 37.72%, which again differs from the Chapter III results because of the removed data.

There are again two possible ways to determine error. They are using 0.5 as the prediction breakpoint to determine accept/reject and finding a breakpoint that minimizes error. The point of minimum error is determined by using an S-plus function that is included in Appendix D. The error for both cases is the sum of the recruits we predicted as completing two years who attrited and recruits we predicted would attrite who did not divided by the number of recruits in the data set. For the 0.5 model the error is 37.44% and the minimum error is 37.36% at the prediction point of 0.51. Neither of these provides much improvement over the naïve model.

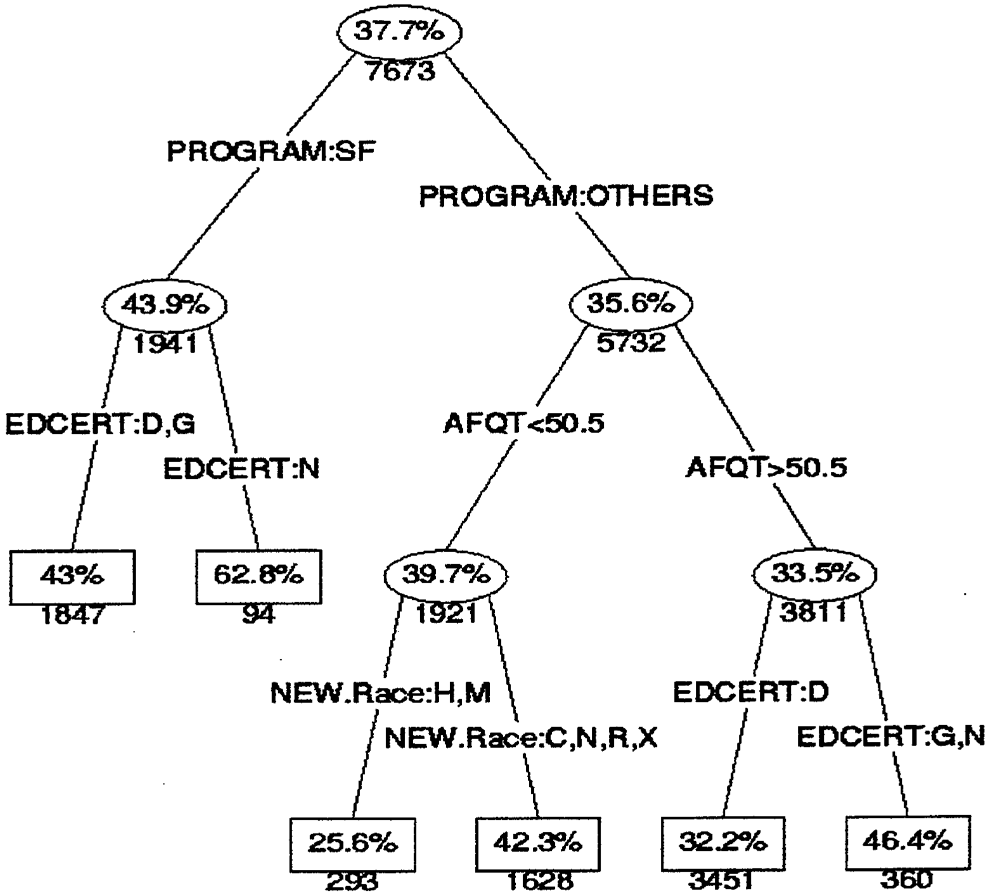
It is also important to look at the recruits that would not have been enlisted based on these models. For the 0.5 model 399 recruits would not have been enlisted. Of these 399 recruits, 189 did not attrite and 210 did, a 47.37% error. For the 0.51 model, 323 recruits would not be enlisted, 148 who did not attrite and 175 who did, a 45.82% error.

The final step is to validate the model. This is completed in the same fashion as for the entire data set model, using the cross-validation procedure. Our goal is to find an error that is near the error of the original model. The cross-validated error was 37.69% for the 0.51 model and 37.65% for the 0.5 model. Neither of these varies largely from the prediction using the entire data set, but the 0.5 model is closer to the model results than the 0.51 model. It was decided that these models were satisfactory for use.

2. Classification Tree

The classification tree was started by modeling Unsuitable Attrition by the same prediction variables that were used for the logistic model using the procedures as detailed in Chapter II. Cross-validation was run on the tree until the minimum deviance was found twice on the same size tree. Size 5 was found twice in three cross-validation runs. However, all of the terminal nodes in the size 5 tree end with a majority of non-attrites. This results in an error identical to that of the naïve model. It was found that a size 6 tree does have a terminal node with a majority of attrites. It was therefore determined to use a tree of size 6 for the model. Figure 4 is the graphical result, where the number below each node shows the number of recruits in that node. The number in the node is the percentage of recruits in that node who are from the "unsuitable attrition" group (yes's). For example, 43.9% of the 1941 recruits in the SF program would unsuitably attrite.

Figure 4: Modified Data Set Classification Tree



This tree has a 37.40% misclassification error and there are 94 recruits that would not be enlisted. 35 of these recruits did not attrite and 59 did, an error of 37.23% for the recruits that would not be enlisted based on the model. The recruits that are not enlisted are recruits with EDCERT code of N (non-grad) from the PROGRAM SF (seafarer). This is considered the final model since cross-validation was used in building the model.

3. Model Results

Neither of the two types of models provides a great improvement in the modified data set. The error rate is basically the same in the classification tree and the regression model. However, the logistic regression excludes more recruits from acceptance and has higher error rates than the classification tree in the recruit exclusion area. It is also evident in both models that EDCERT code of N (non-grad) has the effect of increasing the chance of unsuitable attrition.

C. DATA SET COMPARISONS

It is now the goal to compare the results of the two data sets. It is first obvious that no substantial prediction capability was found in the study of either data set. The best error improvement is in the classification tree for the entire data set, but it was only a 0.52% improvement above the naive model.

It is also noted that in all 4 models, the EDCERT code of N (non-grad) has some type of significance. This is seen as a high loss probability coefficient in the logistic models and with it being combined with other characteristics to predict loss in the tree models. However, the characteristics associated with an unsuccessful non-grad differ between the two data sets. The one other issue of importance that effects all of the models is in the Race/Ethnic code of M (Asian). This code greatly decreases loss probability in the two logistic models, and does not show any splits in any tree branch that increases chance of unsuitable attrition.

D. FURTHER MODEL EXTENSIONS

Since the basic models did not result in prediction capability that greatly improved from the naïve model, three further modeling attempts were undertaken. There was also one modification made using the original models. None of the new attempts greatly improved prediction capabilities, but will be explained for any further use of this study. The three new models were a hybrid tree, splitting unsuitable attrition into two categories, and modeling the large waiver types individually.

A hybrid tree was built by creating a classification tree of size 4 for the entire data set and size 5 for the modified data set. The tree then had a logistic regression run in each of the terminal nodes. The goal of this was to have an easy model for use once the tree was used for initial data separations. However, the results were no better than the actual tree for the entire data set and only 0.07% better than the best model of the modified data set. Since this model did not provide significant improvement, it was decided that the more complicated model was not useful.

In the second new model type, unsuitable attrition was divided into two segments: Code 970 losses (entry-level separation) and all other losses. This was undertaken since entry-level losses make up over 50% of the unsuitable losses in both of the data sets. These two separate segments were then modeled using the same procedures as were used on the initial data sets. The results did not improve the misclassification error for either the entire data set or the modified data set when the two loss segments were combined. However, a couple of important points were found. In the modified data set, it was found that females had a very low unsuitable attrition in the non-entry level segment. This

percentage was 8.96% compared to 18.66% for men in the same category. It was also found that lower AFQT scores were very important in predicting the loss of recruits to entry level separations in both the entire data set and the modified data set. The tree models identified the score of importance to be between 76 and 79 depending on the data set. The difference in attrition percentages between being above and below the critical score was between 6 and 8% depending on the data set.

The third attempt was modeling non-minor misdemeanors and drug-related waivers by themselves. This was modeled for both data sets for each of these two categories. The greatest improvement was 1.3% above the naïve model for any segment. The other three segments showed smaller improvements. It was determined that the improvement was not large enough to justify the models. This procedure is also limited by only being able to use on two waiver types because of the small numbers in the other waiver categories. One point of interest for this model is found in the drug-related waivers. Sex was found to be extremely important in both data sets. In the entire data set, 23% of the females who received drug waivers attrited unsuitably compared to 39% of the males with drug waivers. There was an even larger difference in the modified data set, with females at 24% and males at 42%.

One other procedure was attempted using the logistic models that were created initially for the two data sets. This procedure was to minimize the error associated with only the recruits who would not have been enlisted based on the model. This was done using an S-Plus function that is included in Appendix D. This is justified by arguing that it is more important to see the percentage of recruits that would have succeeded but are not

recruited based on our model than to look at the recruits who are recruited and subsequently attrite. With this analysis, the entire data set breakpoint was 0.63, and the model yielded a 20.00% error. However, in doing this it is found that only a total of 15 recruits would not have been enlisted from the two years of data. For the modified data set, the breakpoint was 0.61, which had a 24.32% error. In the modified data set, 37 recruits would not have been enlisted from the two years of data. This procedure is limited in the fact that it only affects a very small number of recruits and therefore will not provide great improvements to the overall unsuitable attrition percentages.

Although these attempts did not provide major improvements as had been hoped, they did give some extra insights. Therefore, some of these results will be discussed in the conclusions and recommendations.

V. CONCLUSIONS AND RECOMMENDATIONS

A. STUDY CONCLUSIONS

The first goal of this study was to compare unsuitable attrition between recruits with moral waivers and recruits without moral waivers. It was found that these rates are different. In the entire data set recruits with moral waivers (34.02%) had a 9.3 percent higher unsuitable attrition than the recruits without moral waivers (24.70%).

However, it was known at the start that the data set for this study contained errors because program waivers were recorded as moral waivers. To account for this, the data was purged of rates and programs that could receive program waivers to see if there was an impact. It was found in this modified data that recruits with moral waivers (37.26%) had a higher unsuitable attrition than recruits without (26.34%). This is a 9.9 percent higher attrition rate, a slightly larger difference than was seen in the entire data set. It was also found that the attrition rates for the modified data set are higher overall than that of the entire data set.

Unsuitable attrition was also analyzed for each waiver category. In the entire data set, recruits in each of the legitimate waiver categories (legitimate excludes "other", N/A, and fewer than 3 minor misdemeanors) had a higher unsuitable attrition rate than those without moral waivers. The two highest rates in the entire data set were for the two felony waiver categories. In the modified data set, all of the legitimate waiver categories except 3 or more minor misdemeanors had higher unsuitable attrition rates than the rates

of recruits without moral waivers. The two highest rates are for minor traffic waivers and adult felony waivers.

The two data sets do have obvious differences in their unsuitable attrition rates. The rates are higher in the modified data set, which supports the hypothesis that program waivers are affecting the perceived rate of unsuitable attrition for recruits with moral waivers. It appears that the program waiver data recording error is decreasing the unsuitable attrition rate that is being found for recruits with moral waivers. It must also be noted that the drug waiver category decreases in size the most when the data is modified. This is an expected result since most of the program waivers are due to more stringent drug waiver requirements. Within the drug abuse category, the unsuitable attrition rate increases by 7.66% while the number of recruits in this category falls by 2,329.

The next goal of the study was to create models to predict unsuitable attrition among the recruits that require moral waivers. These models were developed for both data sets. In developing the models, the goal was to create a logistic regression model and a classification tree for use with each data set. This created four models, none of which resulted in a large improvement over the naïve model of just using the rate of attrition loss. The best improvement found was 0.52% using the logistic model for the entire data set.

Further model extensions were also undertaken due to the lack of success with the initial models. These were a hybrid tree, splitting data by type of unsuitable attrition, and modeling the large waiver types individually. None of these models increased predictive capability enough to justify their use.

Even though none of the models provided useful prediction models, some interesting point were found. The most significant issue was that a recruit with the non-grad code in the EDCERT column had a higher chance of unsuitable attrition in all four of the initial models. It was also found in all four models that recruits with the race/ethnic code of Asian have a large improvement in their chance of not being lost due to unsuitable attrition. The other important points were in the model extensions. Recruits with AFQT scores below 76 had a substantial increase in the probability of being an entry level separation and males with drug related waivers are much more likely to have unsuitable attrition than females with the same waivers. One other important point must also be included in the results: the fact that none of the prior service recruits had unsuitable attrition, but were not included in the model development because of data errors.

B. COMPARISONS WITH BACKGROUND RESEARCH

The results of this study concur with previous studies in the finding that recruits with moral waivers do have a higher rate of unsuitable attrition than recruits without moral waivers. The rates found by Bohn and Schmitz are similar to the rates found in the entire data set of this study. Etcho's point that "non-graduate of high school" is a significant variable in predicting attrition for Marine Corps recruits with moral waivers was also found to hold for predicting unsuitable attrition for the Navy data used in this study. However, it is noted that none of the models in the background research attempted predicting once their models were developed. Therefore, the capability for prediction of this model can not be compared with the previous models.

The other background article of interest was from Kannapel who asserted that the increase in attrition by moral waiver recruits was due to the increased number of recruits who were being given moral waivers. The percentage of recruits with moral waivers that is found in this study does concur with his findings. However, this study does not support the claim that higher unsuitable attrition among moral waiver recruits is just an effect of more recruits being given moral waivers. This study finds that attrition for recruits with moral waivers is higher than that of recruits without moral waivers, which he does not address as a cause in his article.

The changes that have been implemented in moral waiver policy after the time of the data for this study appear to be positive steps. The waiting period for adverse alcohol/drug adjudication is supported as both of these waiver groups have higher unsuitable attrition than the non-waiver group in both data sets. Therefore, this policy may decrease unsuitable attrition rates, though that can not be confirmed based on this study. It is also believed that program waivers being treated separately is a good step that will help future studies. The modifications of the data in this study appear to support the assertion that program waivers are affecting the results of moral waiver studies.

C. RECOMMENDATIONS

In providing recommendations based on this study, two major issues arise. The first is that none of the predictive models provide a substantial improvement above current policy. The second point is that the use of the predictive models would exclude some recruits who will succeed. With the recruiting problems that are present at the time of this

study, excluding recruits that would succeed does not seem feasible. Therefore, it is not recommended that any changes be undertaken based on this study. However, the models are available so the end users can make the final decision based on the analysis.

Even though the prediction models do not show a lot of improvement, there are some important results of this study. The main result is that among recruits with moral waivers, the chance of unsuitable attrition greatly increases if the recruit is not a high school graduate. Among the non-high school graduates, the increase is largest for the EDCERT code non-grad, but is also substantially larger for the G.E.D. code in EDCERT compared to graduates. This does bring into question the policy at the time of this thesis of allowing more recruits that are non-high school graduates to enlist.

It is recommended that studies be conducted to address the concern of increasing the number of non-high school graduates that are being enlisted. Since the predictive models in this study only address recruits with moral waivers, no broad statements can be made about the possible implications of the increase.

It is also recommended that a study similar to this one be conducted once data is available that does not include program waivers in the moral waiver data. This would allow a study that is not impacted by removing data due to the unknown nature of waivers in certain programs and/or rates.

Finally, although it is not recommended that the results of this study be used in the current recruiting environment, future use is possible. These models could be used or re-evaluated at a future time if the ability to exclude recruits from consideration is a more feasible policy.

APPENDIX A: Civil Waiver Classifications

This appendix contains a list of civil crimes and their classifications. The list is not all-inclusive list but intended to serve as a guide.

1. Minor traffic violations

Blocking or retarding traffic
Careless driving
Crossing yellow line; driving left of center line
Disobeying traffic lights, signs, or signals
Driving on shoulder
Driving uninsured vehicle
Driving with blocked vision
Driving with expired plates or without plates
Driving without license, or suspended/revoked license
Driving without registration or improper registration
Driving wrong way on one-way street
Failure to comply with officer's directives
Failure to have vehicle under control
Failure to keep right or in line
Failure to signal
Failure to submit report following accident
Failure to yield right-of-way
Following to closely
Improper backing
Improper blowing of horn
Improper turn
Invalid, unofficial or no inspection sticker
Leaving key in ignition
License plate improperly or not displayed
Operating overloaded vehicle
Participating in contest of speed (Note 1)
Speeding (Note 1)
Start; Improper or spinning wheels (Note 1)
Zigzagging or weaving in traffic (Note 1)

Note 1: When not considered reckless driving.

2. Minor non-traffic violations/Minor misdemeanors

Abusive language to provoke breach of peace

Assault
Carrying concealed weapon (other than firearm)
Check, worthless, making or uttering with no intent to defraud or deceive
    ($100 or less)
Criminal trespassing
Curfew violation
Damaging road signs
Discharging firearm through carelessness
Disobeying summons
Disorderly/boisterous conduct; creating disturbance
Disturbing peace
Drinking in public
Drunk in public; drunk and disorderly
Dumping refuse near highway
Failure to appear
Fare/toll evasion
Fighting; participating in an affray
Fornication
Illegal betting or gambling
Juvenile non-criminal misconduct; runaway; truant; incorrigible; wayward; beyond
    parental control
Killing domestic animal
Liquor; Unlawful manufacture, sale or possession
Littering
Loitering
Malicious mischief
Minor in possession of alcohol
Nuisance; Committing
Poaching
Possession of cigarettes by minor
Possession of indecent publications or pictures
Possession of drug paraphernalia
Purchasing, possessing or consuming alcohol by minor
Removing property under lien
Removing property from public grounds
Robbing orchard
Shooting from roadway
Simple assault
Trespass to property
Unlawful assembly
Using or wearing unlawful emblem
Vagrancy
Vandalism

Violation of fireworks laws
Violation of fish and game laws

3. Non-minor misdemeanors

Accessory before or after the fact of a misdemeanor
Adultery
Assault consummated by battery
Behind the wheel .08 blood alcohol content or greater
Bigamy
Breaking and entering less than $500
Check, worthless, making or uttering with intent to defraud or deceive
    ($500 or less)
Conspiring to commit misdemeanor
Contributing to the delinquency of a minor
Criminal mischief
Desecration of a grave
Driving while drugged or intoxicated
Failure to stop and render aid after an accident
Indecent exposure
Indecent, insulting or obscene language communicated directly or by telephone
Leaving the scene of an accident
Looting
Negligent homicide
Petty larceny ($500 or less)
Possession and/or use of marijuana/controlled drug
Reckless driving
Resisting arrest
Sex crime related charges
Slander
Stalking
Stolen property; Knowingly receiving ($500 or less)
Suffrage; Interference with
Unlawful carrying of firearms; concealed weapon
Unlawful entry
Unlawful use of long-distance phone lines
Use of telephone to abuse, annoy, harass, or threaten
Using boat without owner's consent
Willfully discharging firearm so as to endanger life
Wrongful appropriation of motor vehicle; joyriding

4. Felonies

Accessory before or after the fact of a felony
Aggravated assault
Arson
Attempt to commit a felony
Breaking and entering with intent to commit a felony
Bribery
Burglary
Carnal knowledge of a female under 16
Cattle rustling
Car jacking
Check, worthless, making or uttering with intent to defraud or deceive (over $500)
Concealing knowledge of a felony
Conspiring to commit a felony
Criminal libel
Extortion
Forgery; knowingly uttering/passing forged instrument
Graft
Grand larceny; embezzlement (value over $500)
Housebreaking
Indecent acts/liberties with child under 16
Indecent assault
Kidnapping; abduction
Mail matters; destroying, obstructing, stealing, etc.
Mail; Depositing obscene or indecent matter
Maiming; disfiguring
Manslaughter
Murder
Narcotics, dangerous drugs, or marijuana; possession or use of
Pandering
Perjury; subornation of perjury
Possession of controlled substance
Public record: Altering, concealing, or destroying
Rape
Riot
Robbery
Sedition; solicitation to commit sedition
Selling or leasing weapons to minors
Sodomy
Stolen Property; Knowingly receiving (value over $500)

Appendix B: Data Descriptions

This appendix provides a description of the codes that require further explanation from the data set.

A. NAVYLOSS:

1. NON ATTRITION DISCHARGE

(BLANK) = STILL ON ACTIVE DUTY
801 = HONORABLE DISCHARGE EXPIRATION ENLISTMENT
802 = HONORABLE DISCHARGE WITHIN 3 MONTHS OF END OF ENLISTMENT
803 = HONORABLE DISCHARGE CONVENIENCE OF GOVERNMENT (COG) EARLY OVER 3 MONTHS TO 12 MONTHS
806 = HONORABLE DISCHARGE COG FROM USNR TO ENLISTED USN
809 = HONORABLE DISCHARGE COG TO ENTER STAR PROGRAM NMPC 1133.30
811 = HONORABLE DISCHARGE COG TO ENTER SCORE PROGRAM NMPC 1440.27
816 = HONORABLE DISCHARGE FULFILLMENT SERVICE UNIVERSAL MILITARY TRAINING (UMT)
833 = HONORABLE DISCHARGE, UNSUITABILITY - HOMOSEXUAL
841 = GENERAL DISCHARGE EXPIRATION ENLISTMENT
842 = GENERAL DISCHARGE COG WITHIN 3 MONTHS OF END OF ENLISTMENT
843 = GENERAL DISCHARGE COG OVER 3 MONTHS TO 12 MONTHS END OF ENLISTMENT
846 = GENERAL DISCHARGE COG FROM USNR TO ENLISTMENT USN
849 = GENERAL DISCHARGE STAR NMPC 1133.13
850 = GENERAL DISCHARGE COG TO ENTER COLLEGE, UNIVERSITY, OR VOCATIONAL SCHOOL
856 = GENERAL DISCHARGE FULFILLMENT UMT SERVICE
931 = RELEASED TO INACTIVE DUTY FLEET RESERVE
932 = RELEASED TO INACTIVE DUTY RETIRED NON-DISABILITY
942 = RELEASED TO INACTIVE DUTY TRANSFERRED TO NAVAL RESERVE
943 = RELEASED INACTIVE DUTY, TEMPORARY ACTIVE DUTY COMPLETED (948)
953 = ENLISTMENT CANCELLED
980 = AWAITING RESULTS OF APPELATE REVIEW

996 = LOSS DATA CORRECTION
997 = ACCOUNTING LOSS NAVY STRENGTH
998 = CANCEL ERRONROUS STRENGTH GAIN
999 = NMPC DISCHARGE, NO DISCHARGE WITHIN 10 YEARS OF LAST EVENT

2. ATTRITION DISCHARGES

804 = HONORABLE DISCHARGE DISABILITY SEVERANCE PAY
805 = HONORABLE DISCHARGE DISABILITY NO SEVERANCE PAY
807 = HONORABLE DISCHARGE COG TO ACCEPT COMMISSION
808 = HONORABLE DISCHARGE COG ACCEPT APPOINTMENT OTHER SERVICE
813 = HONORABLE DISCHARGE COG OTHER (810,945)
814 = HONORABLE DISCHARGE DEPENDENCY OR HARDSHIP
815 = HONORABLE DISCHARGE MINORITY
817 = HONORABLE DISCHARGE UNSUITABILITY INAPTITUDE
818 = HONORABLE DISCHARGE UNSUITABILITY OTHER THAN INAPTITUDE (819,820,821,822,823)
824 = HONORABLE DISCHARGE SECURITY
825 = HONORABLE DISCHARGE UNFITNESS (826,827,828,829)
830 = HONORABLE DISCHARGE GOOD OF SERVICE
831 = HONORABLE DISCHARGE MISCONDUCT (832,833)
832 = HONORABLE DISCHARGE DRUG EXEMPTION PROGRAM
844 = GENERAL DISCHARGE DISABILITY WITH SEVERANCE PAY
845 = GENERAL DISCHARGE DISABILITY NO SEVERANCE PAY
853 = GENERAL DISCHARGE COG OTHER REASONS
854 = GENERAL DISCHARGE DEPENDENCY OR HARDSHIP
855 = GENERAL DISCHARGE MINORITY
857 = GENERAL DISCHARGE UNSUITABILITY INAPTITUDE
858 = GENERAL DISCHARGE UNSUITABILITY OTHER THAN INAPTITUDE (859,860,861,862,863)
864 = GENERAL DISCHARGE SECURITY
865 = GENERAL DISCHARGE UNFITNESS (866,867,868,869)
870 = GENERAL DISCHARGE GOOD OF SERVICE
871 = GENERAL DISCHARGE MISCONDUCT (872,873)
872 = GENERAL DISCHARGE HOMOSEXUAL
873 = GENERAL DISCHARGE DRUG ABUSE OTHER THAN ALCOHOL
881 = UNDESIRABLE DISCHARGE SECURITY
882 = UNDESIRABLE DISCHARGE UNFITNESS (883,884,885,886,887)
887 = UNDESIRABLE DISCHARGE GOOD OF SERVICE
888 = UNDESIRABLE DISCHARGE MISCONDUCT (890)
889 = UNDESIRABLE DISCHARGE AMNESTY

890 = OTHER THAN HONORABLE DISCHARGE
901 = BAD CONDUCT DISCHARGE SPECIAL COURT MARTIAL (903,905)
902 = BAD CONDUCT DISCHARGE GENERAL COURT MARTIAL (GCM) (904,906)
911 = DISHONORABLE DISCHARGE, GENERAL COURT MARTIAL (912,913)
933 = RELEASED TO INACTIVE DUTY RETIRED DISABILITY
944 = RELEASED INACTIVE DUTY, HARDSHIP OR DEPENDENCY
952 = DIED ON ACTIVE DUTY
954 = APPOINTMENT OFFICER STATUS
955 = APPOINTMENT NAVAL AVIATION CADET
956 = APPOINTMENT AVIATION OFFICER CANDIDATE
957 = APPOINTMENT OFFICER CANDIDATE
958 = APPOINTMENT NAVAL ACADEMY MIDSHIPMAN
959 = APPOINTMENT NROTC MIDSHIPMAN
960 = APPOINTMENT OTHER SERVICE ACADEMY
961 = APPOINTMENT NAVAL AVIATION OBSERVER
970 = ENTRY LEVEL SEPARATION
971 = VOID ENLISTMENT
972 = REMOVED FROM ROLLS

B. ATTRITE

1 = attrite within 2 years
0 = did not attrite in 2 years
-1 = not in 2 years as of 30 June 1998

C. PRIOR_SV

0 = no prior service
1,5 = NAVET continuous service
2,3,4,6 = NAVET broken service
7,8,9 = other service veteran

D. SENGRAD

S = Senior
N = Non-grad
G = High school diploma

E. EDCERT

D = Diploma grad
G = G.E.D.
N = Non-grad
P = Senior

F. CIV_CODE

1 = less than HS diploma
7 = correspondence school diploma
8 = one semester of college (for non-grads)
B = adult ed diploma
C = occupational program certificate of attendance
D = associate degree
E = GED or other such test
G = nursing diploma
H = home study
J = hs certificate of attendance
K = baccalaureate degree
L = hs diploma
M = in process of attending college or adult ed (for non-grads)
N = master's degree
R = post master's degree
S = senior
U = doctorate
W = first professional degree

G. RACE

C = Caucasian
N = Black
X = Other
Z = Unknown
R = American Indian
M = Asian

H. ETHNIC

1 = Spanish descent
2 = American Indian
3 = Asian American
4 = Puerto Rican
5 = Filipino
6 = Mexican-American
7 = Eskimo
8 = Aleut
9 = Cuban American
G = Chinese
J = Japanese
K = Korean
S = Latin American Hispanic
D = Indian
V = Vietnamese
E = Melanesian
W = Micronesian
L = Polynesian
Q = Other pacific island
X = Other
Y = None
Z = Unknown

I. PROGRAM

2YO	Two year option
3YO	Three year option
5YO	Five year option
AEF	Advanced Electronic Field
ATF	Advanced Technical Field
DIVR	Diver
JOBS	Jobs Program
NF	Nuclear Field
SF	Sea Farer
SG	School Guarantee
TEP	Temporary Active Reserve Enlisted Program

J. RATE

ABE	Aviation Boatswain's Mate - Launching & Recovery Equipment
ABF	Aviation Boatswain's Mate - Fuels
ABH	Aviation Boatswain's Mate - Aircraft Handling
AC	Air Traffic Controller
AD	Aviation Machinist's Mate
AE	Aviation Electrician's Mate
AECF	Advanced Electronics Career Field
AG	Aerographer's Mate
AIRC	Aircrew - Rescue Swimmer
AIRR	Aircrew - Non-Rescue Swimmer
AK	Aviation Storekeeper
AME	Aviation Structural Mechanic – Safety Equipment
AMH	Aviation Structural Mechanic - Hydraulics
AMS	Aviation Structural Mechanic - Structures
AN	Airman
AO	Aviation Ordnanceman
AS	Aviation Support Equipment Technician
AT	Aviation Electronics Technician
AW	Aviation Warfare Systems Operator
AZ	Aviation Maintenance Administration
BT	Boiler Technician
BU	Builder
CE	Construction Electrician
CM	Construction Mechanic
CTA	Cryptologic Technician Admin
CTI	Cryptologic Technician Interpretive
CTM	Cryptologic Technician Maintenance
CTO	Cryptologic Technician Communications
CTR	Cryptologic Technician Collection
CTT	Cryptologic Technician Technical
DC	Damage Controllman
DIVE	Diver
DK	Disbursing Clerk
DP	Data Processing Technician
DS	Data Systems Technician
DT	Dental Technician
EA	Engineering Aid
EM	Electrician's Mate
EN	Engineman
EO	Equipment Operator
EOD	Explosive Ordnance Disposal

ET	Electronics Technician
ETS	Electronics Technician - Submarine
EW	Electronics Warfare Technician
FC	Firecontrolman
FN	Fireman
GM	Gunner's Mate
GSE	Gas Turbine Systems Technician - Electrical
GSM	Gas Turbine Systems Technician - Mechanical
HM	Hospital Corpsman
HT	Hull Technician
IC	Interior Communications Electrician
IM	Instrumentman
IS	Intelligence Specialist
JO	Journalist
LI	Lithographer
ML	Molder
MM	Machinist's Mate
MMS	Machinist's Mate
MN	Mineman
MR	Machinery Repairman
MS	Mess Management Specialist
MSS	Mess Management Specialist (Sub)
MT	Missile Technician
NF	Nuclear Field
OM	Opticalman
OS	Operations Specialist
PC	Postal Clerk
PH	Photographer's Mate
PM	Patternmaker
PN	Personnelman
PR	Aircrew Survival Equipmentman
QM	Quarter Master
RM	Radioman
RP	Religious Program Specialist
SH	Ship's Serviceman
SK	Storekeeper
SKS	Storekeeper - Submarine
SM	Signalman
SN	Seaman
SPEC	Special Warfare
STG	Sonar Technician - Surface
STS	Sonar Technician - Submarine
SUB	Submarine School

SW	Steelworker
TM	Torpedoman's Mate
TMS	Torpedoman's Mate - Submarine
UT	Utilitiesman
YN	Yeoman
YNS	Yeoman – Submarine

K. ACC_WAIV

1. First character

A	age
B	dependents
C	mental qual
D	moral qual
E	previous DQ separation-reenlistment code
F	time lost on prior enlistment
G	last separated because existed prior to service
H	medical qual
J	sole survivor restrictions
K	education qual
L	alien status
M	refused to sign loyalty certificate
N	conscientious objector
P	prior service pay-grade
Q	skill(s) requirement
X	not elsewhere classified
Y	not applicable

2. Second character (explanation for moral waiver, Y if first character not D)

A	minor traffic offense
B	< 3 minor misdemeanor
C	>=3 minor misdemeanor
D	non minor misdemeanor
E	felony (adult)
F	felony (juvenile)
G	pre-service drug abuse
H	pre-service alcohol abuse
X	other
Y	N/A

3. Third character (authority level)

A	Navy Department
B	Commander, Navy Recruiting Command
D	Commanding Officer, NRD
E	Commander, Navy Recruiting Area
Y	N/A

L. NRD

1. East Region

102	New England
103	Buffalo
104	New York
118	Columbus
119	Philadelphia
120	Pittsburgh
122	Michigan

2. South Region

310	Montgomery
312	Jacksonville
313	Atlanta
314	Nashville
315	Raleigh
316	Richmond
334	New Orleans
348	Miami

3. North Region

521	Chicago
527	Kansas City
528	Minneapolis
529	Omaha
531	Dallas
532	Houston
542	Indianapolis
547	St. Louis

4. West Region

825	Denver
830	Albuquerque
836	Los Angeles
837	Portland
838	San Francisco
839	Seattle
840	San Diego
846	San Antonio

Appendix C: Sample Data

This appendix contains a sample of the data used for this study. SSN and DOB have been omitted to protect privacy.

RESDT_TT	CANDATE	DEPDAYS	LOSSDATE	NAVY LOSS	SERVDAYS	ATTRITE
5/25/94	11/1/94	160				0
2/16/95	3/9/95	21	Apr 96	871	389	1
8/2/94	10/4/94	63	Dec 94	970	58	1
7/21/95	8/16/95	26				0
10/28/94	11/29/94	32				0
	8/2/95					0
12/29/94	8/23/95	237				0
1/10/95	2/27/95	48				0
8/7/95	9/20/95	44				0
11/29/94	2/23/95	86				0
6/14/94	11/22/94	161				0
12/29/95	1/3/96	5	Dec 96	887	333	1
12/12/95	12/28/95	16	May 98	871	855	0
3/23/95	9/20/95	181				0
11/29/94	12/8/94	9				0
5/30/96	8/29/96	91				-1
10/27/94	9/6/95	314				0
9/26/95	3/12/96	168	Oct 96	813	203	1
11/30/95	2/14/96	76				0
3/28/96	4/22/96	25	Mar 98	888	678	0
10/26/94	1/25/95	91				0
7/12/94	7/6/95	359				0
6/20/95	6/11/96	357				0
8/12/96	8/12/96	0				-1
2/25/94	11/7/94	255				0
5/30/96	8/22/96	84				-1
6/28/94	11/2/94	127				0
5/24/95	11/6/95	166	Jun 96	853	208	1
10/24/95	10/30/95	6				0
5/30/95	11/29/95	183				0
7/6/95	8/15/95	40				0
1/13/95	4/10/95	87				0
8/26/94	11/3/94	69	Feb 98	871	1186	0

PRIOR_SV	AFQT	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	SENGRAD
0	49	49	48	57	49	55	53	53	45	46	41	G
0	53	39	52	55	50	57	52	49	48	43	51	G
0	71	50	63	54	53	62	63	57	54	53	62	G
0	81	60	63	60	55	54	53	69	55	68	61	G
0	78	58	58	57	59	62	65	42	58	52	58	G
2	0	0	0	0	0	0	0	0	0	0	0	
0	72	67	59	53	53	55	58	61	61	65	62	S
0	84	58	65	54	58	59	58	61	63	57	46	G
0	50	49	48	55	49	52	48	67	50	58	66	G
0	94	62	64	60	55	62	61	56	67	58	66	G
0	89	54	65	58	57	62	64	47	63	59	62	G
0	84	56	57	57	58	58	54	57	66	63	55	N
0	62	47	48	60	53	58	52	47	52	57	53	G
0	82	67	57	58	58	53	55	61	64	61	61	S
0	39	39	45	43	58	51	62	36	53	45	34	G
0	77	60	57	57	55	50	46	65	60	62	66	G
0	41	61	44	52	55	57	52	63	44	46	63	S
0	43	58	41	55	52	58	54	64	48	59	59	G
0	35	43	41	55	42	45	47	47	45	40	36	G
0	36	42	42	52	56	49	55	45	41	46	51	G
0	85	62	59	58	55	60	57	49	67	59	51	G
0	53	50	48	54	53	55	59	49	52	49	41	S
0	81	55	64	56	50	62	51	44	64	53	56	S
8	80	63	59	61	58	54	63	53	53	64	61	G
0	96	60	66	60	55	62	65	51	67	65	62	G
0	39	43	46	45	53	62	60	49	52	47	51	G
0	41	43	44	52	41	49	53	41	52	54	46	G
0	53	51	48	54	45	55	51	62	57	50	56	G
0	89	63	63	59	54	62	68	61	65	68	55	G
0	56	35	56	48	56	61	47	53	53	47	43	G
0	57	54	57	54	49	51	47	41	51	44	36	G
0	50	43	51	45	50	59	60	49	61	58	53	G
0	50	54	41	54	58	61	62	47	53	40	43	G

EDYRS	EDCERT	CIV_CODE	SEX	RACE	ETHNIC	PAY_TT	PAYGRADE
12	D	L	M	C	Y	4	1
12	D	L	M	N	Y	2	1
12	D	L	M	C	Y	1	1
12	D	L	M	C	Y	4	1
12	D	L	F	C	Y	4	1
12	D	L	M	N	Y		4
12	D	L	M	C	Y	4	1
12	D	L	M	C	Y	5	1
12	D	L	M	C	Y	3	1
15	D	D	M	C	Y	4	3
12	D	L	M	C	Y	5	1
11	G	E	M	C	Y	2	1
12	D	L	M	C	Y	3	1
12	D	L	M	C	Y	4	1
12	D	L	F	N	Y	3	1
12	D	L	M	C	Y	4	1
12	D	L	M	C	Y	4	1
12	D	L	M	C	Y	1	1
10	D	B	M	N	Y	3	1
12	D	L	M	C	Y	2	1
12	D	L	M	C	6	4	1
12	D	L	M	N	Y	4	1
12	D	L	M	C	Y	4	1
12	D	L	M	C	Y	4	3
12	D	L	M	C	Y	5	1
12	D	L	M	N	Y	3	1
12	D	L	F	N	Y	3	1
12	D	L	M	C	Y	1	1
12	D	L	M	C	Y	3	1
9	D	8	M	N	Y	4	1
12	D	L	M	C	Y	3	1
12	D	L	M	C	1	4	1
12	D	L	F	C	Y	4	1

PROGRAM	RATE	TERM	DEPEND	ACC_WAIV	NRD
SG	YN	4	0	YYY	103
SG	AO	4	0	YYY	104
SG	MSS	4	0	XYD	104
3YO	FN	3	0	YYY	103
AEF	ET	6	0	YYY	119
SF	GSM	4	0	YYY	103
5YO	AIR	5	0	DGD	120
SG	YN	4	0	YYY	103
SF	AN	4	0	YYY	120
NF	MM	6	0	YYY	103
NF		6	0	KYA	119
SG	AE	4	0	YYY	120
SF	FN	4	0	DDD	120
AEF	ET	6	0	YYY	120
3YO	SN	3	0	YYY	119
AEF	AEC	6	0	YYY	103
SG	MM	4	0	YYY	103
3YO	AN	3	0	YYY	521
SG	MS	4	1	BYD	521
SF	FN	4	0	YYY	521
NF		6	0	YYY	521
SG	OS	4	0	YYY	521
NF		6	0	YYY	521
SG	AMS	4	2	BYD	838
NF		6	0	YYY	547
3YO	SN	3	0	YYY	521
SF	SN	4	0	YYY	846
SG	IC	4	0	DDD	547
SF	SN	4	0	YYY	547
SG	AZ	4	0	YYY	310
SF	AN	4	0	YYY	531
SG	ABH	4	0	YYY	104
TEP	AO	4	0	YYY	314

Appendix D: S-Plus Functions

This appendix contains the functions that were written for use with this thesis in the data analysis. All of these functions are written for use in S-Plus®.

1. Function to convert N/A to zero

```
function(frame)
{
# This function takes a data frame and breaks it apart to look for
# N/A values.  When N/A values are found, it sets them to 0 if they are
# from a matrix or a numeric vector.  If the N/A values are from a
# factor, the factor with missing data is replaced by a new factor with
# one more level, labeled "NA", which records the missing data. Ordered
# factors are treated similarly, except the result is an unordered
# factor. If frame is a model frame, the response variable can be
# identified.  Any rows for which the response is missing are removed
# entirely from the model frame.  This function is a modification of the
# S-plus function na.gam.replace, which would set the N/A values to the
# mean. Individual vectors from a data frame can be passed into this
# function to change that particular column.
   vars <- names(frame)
   if(!is.null(resp <- attr(attr(frame, "terms"), "response"))){
      vars <- vars[ - resp]
      x <- frame[[resp]]
      pos <- is.na(x)
      if(any(pos)) {
         frame <- frame[!pos,  , drop = F]
         warning(paste(sum(pos), "observations omitted due to missing
           values in the response"))
      }
   }
   for(j in vars) {
      x <- frame[[j]]
      pos <- is.na(x)
      if(any(pos)) {
         if(length(levels(x))) {
            xx <- as.character(x)
            xx[pos] <- "NA"
            x <- factor(xx, exclude = NULL)
         }
         else if(is.matrix(x)) {
            ats <- attributes(x)
            w <- !pos
            x[pos] <- 0
            attributes(x) <- ats
         }
         else {
            ats <- attributes(x)
            x[pos] <- 0
            attributes(x) <- ats
         }
         frame[[j]] <- x
```

```
      }
   }
   frame
}
```

2. Minimize total model error

```
function(obj)
{
# This function finds the minimum error of a two-way table from the
# predict function and a response variable of a glm model.  It requires
# one of the responses in the response variable to be "yes."  The model
# calls the user-defined function first.occurrence in its operation.
# The input is an object of class glm.  The output is the best point
# which can be accessed with $p and the error which is accessed with $r.
   call.strs <- as.character(obj$call)
   if(class(obj)[1] == "glm") {
      data.name <- call.strs[4]
   }
   else {
      stop("Only use on glm models")
   }
   data.location <- find(data.name)
   if(length(data.location) == 0) {
      stop("Can't find data set")
   }
   data <- get(data.name, where = data.location[1])
   tilde <- first.occurrence(call.strs[2], "~")
   if(tilde == 0) {
      stop("Can't find response")
   }
   resp.name <- substring(call.strs[2], 1, tilde - 2)
   best <- 1
   yes <- data[, resp.name] == "yes"
   for(i in 1:100) {
      point <- i * 0.01
      guess <- predict(obj, type = "response")
      thetab <- table(guess > point, yes)
      if(thetab[1, 1] + thetab[1, 2] == sum(thetab)) {
         error <- thetab[1, 2]/sum(thetab)
      }
      else {
         error <- (thetab[1, 2] + thetab[2, 1])/sum(thetab)
      }
      if(error < best) {
         best <- error
         bestpt <- point
      }
   }
   results <- list(p = bestpt, r = best)
   results
}
```

3. Cross-Validation

```
function(obj, n = 10, verbose = F, seed, threshold = 0.5)
{
#
# xval: Function to do cross-validation
#
# This function takes in a fitted model and cross-validates
# in on the data originally used (if it can find it). It does
# this by generating a permutation of the numbers from 1 to the
# number of data points (finding the data by grabbing its name
# from the call and using find() and get(), so it won't work if
# no "data=" was specified and it will be fooled if the data
# is different now than it was at the time the model was created).
# Then it breaks the data into n (default: 10) parts and uses
# the subset= argument to run the model n times with each part
# left out in turn. It accumulates the RSS's from each of these
# n models in an lm model (or misclassification error, in the case
# of a glm model) and reports the total.
#
# This function originally written by Professor Sam Buttrey of Naval
# Postgraduate School, Monterey, CA for OA 3104 class.  Original
# function modified slightly for use in thesis.
#
# Arguments: obj: fitted model object
#              n: number of pieces to use (default: 10)
#        verbose: logical: if TRUE, print info for each call
#           seed: if supplied, use this in a call to set.seed()
#                 to initialize the random number generator
#      threshold: threshold value for predictions in glm model
#
# Return value: cross-validated RSE for lm
#               cross-validated misclassification error for glm
#
# Extract the call from the object, convert to character. If you
# "deparse" first, you get the whole call back. If you just convert,
# it breaks it all up, and the name of the data set, if there is
# one, is in the third position. This can be fooled! Does not
# deal with the case where the call already has a subset.
#
   old.call <- paste(as.character(deparse(obj$call)), collapse ="")
   call.strs <- as.character(obj$call)
   if(class(obj)[1] == "lm")
      data.name <- call.strs[3]
   else if(class(obj)[1] == "glm")
      data.name <- call.strs[4]
   else stop("Sorry, only glm and lm models are supported.")
   if(length(grep("*subset", old.call)) > 0)
      stop("Right now I'm not going to handle this case!")
   if(length(data.name) == 0) {
      stop("This function requires an lm object created with
        \"data=\"\n")
   }
#
#
# Find the data and get it. Oh, and count its rows.
```

```
#
   wheres.the.data <- find(data.name)
   if(length(wheres.the.data) == 0)
      stop(paste("Can't find data set", data.name, "\n"))
   data <- get(data.name, where = wheres.the.data[1])
   nrow.data <- nrow(data)
#
# One more thing we'll need is the name of the response. This is in the
# second position, up to the tilde. It's handy to use a function to
# extract this.  The function takes everything up to the second
# character before the tilde. (The first char. before the tilde is a
# space.)
#
   tilde <- first.occurrence(call.strs[2], "~")
   if(tilde == 0)
      stop("Can't find name of response. Weird.")
   resp.name <- substring(call.strs[2], 1, tilde - 2)
#
# Make "response" be the numeric vector of responses. Handle the
# case where "resp.name" isn't a column of the data frame (maybe
# it's a function of a column)
#
   if(!any(names(data) == resp.name)) {
      attach(data, pos = 1)
      response <- eval(parse(text = resp.name))
      detach(1)
   }
   else response <- data[, resp.name]
   if(!missing(seed))
      set.seed(seed)
   samp <- sample(nrow.data)
   chunk.start <- round(seq(1, nrow.data, len = n + 1))[ - (n + 1)]
   rss.total <- 0
   misclass.total <- 0
#
# Now the big loop. For each chunk, get the chunk, that is, the set of
# row numbers to be excluded on this iteration. Generally that set will
# go from one entry of chunk.start to the next; the last is a special
# case.
#
   for(i in 1:n) {
      if(i == n)
         chunk <- samp[(chunk.start[i]):nrow.data]
      else chunk <- samp[(chunk.start[i]):(chunk.start[i + 1] - 1)]
      assign("chunk", chunk, frame = 1)
#
# The new call (which is a text string) looks just like the old, only we
# add "subset = -chunk" at the end. Then the "eval" line actually runs
# that command.
#
      new.call <- paste(substring(old.call, 1, nchar(
      old.call) - 1), ", subset = -chunk)")
      out <- eval(parse(text = new.call))
#
# Predict on the missing part; accumulate the rss or whatever we're
# using.
```

```
#
      if(class(out)[1] == "glm") {
#
# For glm (that is, logit), let's use misclassification error. Do
# predictions with type = "response"; compare that to the threshold;
# build the (mis)classification table; zero the diagonals (those are
# the correct classifications) and accumulate the rest.
#
         rss.total <- rss.total + out$deviance
         pred <- predict(out, data[chunk,  ], type = "response")
         classif <- table(pred > threshold, data[chunk, resp.name])
         classif[row(classif) == col(classif)] <- 0
         misclass.total <- misclass.total + sum(classif)
      }
      else {
         pred <- predict(out, data[chunk,  ])
         rss.total <- rss.total + sum((pred - response[chunk])^2)
      }
      if(verbose)
         cat("Call ", i, new.call, " gave cum rss ", rss.total, "\n")
   }
#
# We're done. For a glm, we return the misclassification error.  This is
# the total misclassifications divided by nrow (data).  Or for an lm,
# the return value will be the overall RSE.  This is the sqrt of the
# aggregated RSS, divided by nrow (data). That's because each data point
# contributes exactly one squared error.
#
   if(class(out)[1] == "glm")
      return(misclass.total/nrow.data)
   else return(sqrt(rss.total/nrow.data))
}
```

4. Minimize non-acceptance error

```
function(obj)
{
# This function finds the minimum error of the second row of a table
# from the predict function and a response variable of a glm model.
# It requires one of the responses in the response variable to be
# "yes."  The model calls the user-defined function first.occurrence
# in its operation.  The input is an object of class glm.  The output
# is
# the best point which can be accessed with $p and the error which is
# accessed with $r.
   call.strs <- as.character(obj$call)
   if(class(obj)[1] == "glm"){
      data.name <- call.strs[4]
   }
   else {stop("Only use on glm models")}
   data.location <- find(data.name)
   if(length(data.location) == 0){
      stop("Can't find data set")
   }
```

```
   data <- get(data.name, where = data.location[1])
   tilde <- first.occurrence(call.strs[2], "~")
   if(tilde == 0) {
      stop("Can't find response")
   }
   resp.name <- substring(call.strs[2], 1, tilde - 2)
   best <- 1
   yes <- data[, resp.name] == "yes"
   for(i in 1:100) {
      point <- i * 0.01
      guess <- predict(obj, type = "response")
      thetab <- table(guess > point, yes)
      if(thetab[1, 1] + thetab[1, 2] == sum(thetab)) {
         error <- 0
      }
      else {
         error <- thetab[2, 1]/(thetab[2, 1] + thetab[2,2])
      }
      if(error < best && error != 0) {
         best <- error
         bestpt <- point
      }
   }
   results <- list(p = bestpt, r = best)
   results
}
```

5. First-Occurrence function

```
function(string,character)
{
# This is a function that finds the number the first occurrence of the
# "character" in the "string."  It returns the value of the position
# where "character" is found.  This function was originally written by
# Professor Sam Buttrey for OA 3104 taught at Naval Postgraduate School
# in Monterey, CA as a function embedded in another procedure.

   all.chars<-substring(string, 1:nchar(string), 1:nchar(string))
   first<-(1:nchar(string))[all.chars==character][1]
   return(ifelse(length(first)==0,0,first))
}
```

LIST OF REFERENCES

Bohn, Don, Evaluation of the Navy's Moral Waiver Policy: A Case Study of the USS Eisenhower, Commander Navy Recruiting Command, Arlington Va., January 1998.

Bohn, Don and Schmitz, Edward, Waiver Policy and Attrition, Commander Navy Recruiting Command, Arlington Va., June 1996.

Chambers, John M. and Hastie, Trevor J., Statistical Models in S, Pacific Grove, Ca.: Wadsworth and Brooks, Inc., 1992.

Commander Navy Recruiting Command Letter 1130 Ser 341/00950 Dated 21 Sep 98, "Enlisted Policy-Gram # 27," Subj: Alcohol, Drugs, Criminal, and Seal Challenge Program Policy Revisions.

Commander Navy Recruiting Command Letter 1130 Ser 34/01079 Dated 3 Nov 98, "Enlisted Policy-Gram # 01-99," Subj: Area Function and Beers Processing Revisions.

COMNAVCRUITCOMINST 1130.8D with Change 31, Arlington, Va., Commander Navy Recruiting Command, 3 Apr 1992.

Connor, Jeffrey W., The Effects of Pre-Service Criminal History on Recruit Performance in the U. S. Navy, Masters Thesis in Management, Naval Postgraduate School, Monterey Ca., March 1997.

Devore, Jay L., Probability and Statistics for Engineering and the Sciences, Pacific Grove, Ca.: Brooks/Cole Publishing Company, 1995.

Etcho, Leonard L., The Effect of Moral Waivers on First-Term Attrition in the Marine Corps, Masters Thesis in Management, Naval Postgraduate School, Monterey Ca., March 1996.

Hamilton, Lawrence C., Regression with Graphics: A Second Course in Applied Statistics, Belmont, Ca., Duxbury Press, 1992.

Kannapel, Carl, "The Trend in Moral Waivers. Should the Fleet Be Concerned? No," Recruiting Facts, No. 98-03, Commander Navy Recruiting Command, Arlington Va., October 1998.

Venables, W.N. and Ripley, B.D., Modern Applied Statistics with S-Plus, New York: Springer-Verlag, Inc., 1994.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center..2
   8725 John J. Kingman Road, STE 0994
   Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library...2
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, CA 93943-5101

3. Commander, Naval Surface Force...1
   U. S. Atlantic Fleet
   1430 Mitscher Ave
   Norfolk, VA 23551-2494

4. Commander, Naval Surface Force...1
   U. S. Pacific Fleet
   2841 Rendova Road
   San Diego, CA 92155-5490

5. Commander, Navy Recruiting Command..1
   Code 352
   5720 Integrity Drive, Bldg 784
   Millington, TN 38054-5057

6. Professor Samuel E. Buttrey...1
   Code OR/Sb
   Naval Postgraduate School
   Monterey, CA 93943-5002

7. Professor Robert A. Koyak...1
   Code OR/Kr
   Naval Postgraduate School
   Monterey, CA 93943-5002

8. Lieutenant Lyle D. Hall..2
   % Erv R. Hall
   Rt. 1, Box 1273
   Cement, OK 73017